Monitoring crop phenology with street-level imagery using computer vision

Raphaël d’Andrimont, 1,* Momchil Yordanov 1,* Laura Martinez-Sanchez 1, Marijn van der Velde 1

1 European Commission, Joint Research Centre (JRC), Ispra, Italy
* These authors contributed equally to this work.

Abstract

Street-level imagery holds a significant potential to scale-up in-situ data collection. This is enabled by combining the use of cheap high quality cameras with recent advances in deep learning compute solutions to derive relevant thematic information. We present a framework to collect and extract crop type and phenological information from street level imagery using computer vision. During the 2018 growing season, high definition pictures were captured with side-looking action cameras in the Flevoland province of the Netherlands. Each month from March to October, a fixed 200-km route was surveyed collecting one picture per second resulting in a total of 400,000 geo-tagged pictures. At 220 specific parcel locations detailed on the spot crop phenology observations were recorded for 17 crop types (including bare soil, green manure, and tulips): bare soil, carrots, green manure, grassland, grass seeds, maize, onion, potato, summer barley, sugar beet, spring cereals, spring wheat, tulips, vegetables, winter barley, winter cereals and winter wheat. Furthermore, the time span included specific pre-emergence parcel stages, such as differently cultivated bare soil for spring and summer crops as well as post-harvest cultivation practices, e.g. green manuring and catch crops. Classification was done using TensorFlow with a well-known image recognition model, based on transfer learning with convolutional neural network (MobileNet). A hypertuning methodology was developed to obtain the best performing model among 160 models. This best model was applied on an independent inference set discriminating crop type with a Macro F1 score of 88.1% and main phenological stage at 86.9% at the parcel level. Potential and caveats of the approach along with practical considerations for implementation and improvement are discussed. The proposed framework speeds up high quality in-situ data collection and suggests avenues for massive data collection via automated classification using computer vision.

Keywords: phenology, plant recognition, agriculture, computer vision, deep learning, remote sensing, CNN, BBCH, crop type, street view imagery, survey, in-situ, Earth Observation, parcel, in situ

1. Introduction

Street Level Imagery (SLI) provides a human-like view with a detailed perspective on the environment around us. SLI has become ubiquitous with the growth of capturing devices, personalized mapping services (e.g. Google Street View), social media platforms (e.g. Mapillary), and visual positioning systems for autonomous vehicles. Socio-economic, political (Gebru et al., 2017), and environmental (Seiferling et al., 2017) applications have been developed by extracting semantic information from SLI using computer vision algorithms. In combination with overhead imagery, e.g. from satellites or drones, or other image sources (e.g. LiDAR, social media), data integration enables further automated cross-referencing and on-the-fly analysis. SLI is relatively cheap and can be rapidly collected, mapped, and visualized. Various applications are built to take advantage of this, e.g. for rapid damage assessments or inventories after disasters (Lemoine et al., 2013), to check and/or confirm geo-located data previously obtained through e.g. remote sensing (d’Andrimont et al., 2018), or to purposely collect new in-situ data to improve remotely sensed based algorithms.

Smartly surveyed SLI holds a large potential to efficiently source significant sets of in-situ data. Roadside sampling is a viable source of training data for cropland mapping (Waldner et al., 2019). In a previous study on grassland monitoring, SLI were used to evaluate Sentinel-1 and -2 based markers indicative for crop type and bare soil targeting specific parcels (d’Andrimont et al., 2018). Similarly, field pictures have been used in many studies to obtain in-situ information on vegetation, land cover and use, and landscapes. For instance, in-situ repeat photographic imagery has been applied to evaluate the phenological development of natural ecosystems (Sonnentag et al., 2012; Klosterman et al., 2014; Keenan et al., 2014; Nijland et al., 2016). Crowdshared pictures on Flickr have been harnessed for biodiversity monitoring (Stafford et al., 2010; Barve, 2014; ElQadi et al., 2017). Usually, classification of vegetation and crop type or phenology on the imagery relies on human interpretation (e.g. Deus et al. (2016); d’Andrimont et al. (2018); Hufkens et al. (2019); d’Andrimont et al. (2020a)).

Applying deep learning for image recognition is massively speeding up this process. Deep learning is a machine learning method providing a hierarchical representation of a data set by means of various convolutions (LeCun et al., 1995). This method has gained popularity for image processing and data analysis. This is primarily because of its feature learning ability with the automatic extraction of features from raw data - with features from higher levels of hierarchy being formed by the composition of lower level features (LeCun et al., 2015a).
The large learning capacity of deep learning algorithms makes them ideal to perform classification and prediction tasks for a variety of highly complex data (Pan et al., 2010). Different types of architectures were developed for deep learning networks. Convolutional Neural Networks (CNN) are most commonly applied to analyze visual imagery. These are designed to recognize visual patterns directly from pixel images with minimal pre-processing, recognizing patterns with extreme variability and with remarkable robustness to distortions and simple geometric transformations (LeCun et al., 2015b).

Deep learning on imagery has several applications in the agricultural domain (Kamilaris & Prenafeta-Boldú, 2018), including automated robotic weeding (Champ et al., 2020), crop disease recognition (Mohanty et al., 2016), paddy crop stress detection (Anami et al., 2020), phenotyping (Namin et al. (2018)), and precision agriculture (Zheng et al., 2019). Other relevant applications include automatic information extraction from phenocam network pictures (Cao et al., 2021). Yalcin (2017) used a CNN to classify five different crop types on images captured by field cameras from the Turkish agriculture monitoring network. Hufkens et al. (2019) applied image processing techniques on smartphone camera photographs to derive crop phenological parameters. At the same time, large amounts of plant species occurrence data are being collected with citizen science applications such as Pl@ntNet (Götäu et al., 2018) that provide plant species recognition capability.

While promising, only a few studies have used SLI in combination with deep learning to recognize crops. Paliyam et al. (2021) proposed a framework for obtaining geo-referenced crop type labels obtained from vehicle-mounted cameras and tested in Kenya for maize and of sugarcane. Wu et al. (2021) collected a specific SLI dataset for crops, and evaluated several deep CNNs for the classification task. Yan & Ryu (2021) explored Google Street View with deep learning for crop type mapping, indicating that such methods can be efficient and cost-effective. Despite these previous studies, using SLI surveyed combined with deep learning for crop and phenology capturing has not yet been used at its best potential.

1.1. Objectives

In this paper we describe a fully open and operational approach that starts by ground surveying and collecting the SLI, proceeds with classifying crops and crop phenological stages using computer vision, and ends by generating parcel level in-situ data. The objectives are the following:

- To design and test a cheap approach to efficiently collect high quality ground truth for crop types and crop phenological stages
- To collect and to share a representative set of side-looking SLI
- To produce a generic deep learning hyperparameterization computer vision framework tailored to street view imagery for crop type and phenology
- To evaluate to what extent majority voting at parcel level improves the accuracy of the crop type and phenology classification
- To publish all code and training data fully open in one coherent framework (see section 2.5)

2. Materials and Methods

The Materials and Methods section describes 1) the study area, 2) the field data collection including SLI collection, crop and phenology in-situ observations, along with the geo-spatial data-processing to associate the imagery to parcels, 3) the deep learning framework, 4) the accuracy assessment for crops and phenology stage at picture and parcel level and, finally, 5) the code and computing infrastructure.

2.1. Study area

The study area is located within the Dutch province of Flevoland where agriculture is the dominant land use. The region is one of the most productive ones worldwide with yields close to potential (Schaap et al., 2011). Flevoland has ideal features for SLI surveying such as a good road network, a large parcel density and crop type heterogeneity. The extent of the study area is shown in Fig. 1, along with the roads surveyed, the parcels and the locations of the collected ground-truth observations. The route followed about 200 km of predominantly inner roads and was surveyed with a 50 km/h speed limit each month during the vegetation growing season. The surveyed track and driving speed were chosen based on the ability to guarantee a relatively consistent image quality, coupled with the ability to stop safely for detailed ground observations. The survey was designed to ensure a statistically meaningful sample density for the main crop types and their phenological dynamics in the area.

2.2. Field Data Collection

A total of eight survey field campaigns were carried out each month from March to October 2018. The surveyors followed the same track during every campaign and collected roughly the same number of pictures - on average around 49,000 SLI with GPS location information, and 220 in-situ phenology observations for a total of 17 observed different crop types and their characteristic growth stages.

2.2.1. Street-level Imagery acquisition

To collect side-looking SLI, two actions cameras (Sony HDR-AS300R with 128GB miniSD-cards) were mounted on the roof of the car (Delkin Fat Gekko Camera with a suction cup) as shown in Fig. 2. The cameras were equipped with a wide-angle lens (ZEISS Tessar) and were mounted on both (right and left) sides of the roof. The image collection rate was set to one image/second. The lens was chosen based on its ability to capture much of the scene without overly distorting it. The added contrast enhances the difference between the background and the foreground. Two additional reflex cameras
Fig. 1: The study area is in the province of Flevoland in the Netherlands. The specific study area includes the Oost Flevoland and Noord-Oost Polder districts. During 8 surveys a total of 1,691 in-situ crop phenology observations were made at 220 parcel locations alongside the 200-km surveyed track. Projection of the map: Amersfoort / RD NEW (EPSG:28992).

During surveys close-up high-quality detailed images were collected from the side view in both right and left looking directions with about 20% of the image taken up by the sky and another 20-40% taken up by road (depending on the distance between the car and the edge of the road). The remaining portion of the image contains the parcel with the relevant crop.

A total of 199,375 SLI was collected from the roof-mounted action cameras. A breakdown of the image collection per survey date and image usability (i.e. cropland pictures and pictures for which we have ground truth) is listed in Fig. 6.

2.2.2. Phenology Observations

For every campaign there were on average 220 ground-truth observations with a minimum sample of 20 parcels per main crop type. The main crops for which ground-truth were collected are: Carrot (CAR), Green Manure (GMA), Grassland (GRA), Grass Seeds (GRS), Maize (MAI), Onion (ONI), Potato (POT), Summer Barley (SBA), Sugar Beet (SBT), Spring Cereals (SCR), Spring Wheat (SWH), Tulip (TLP), Vegetable (VEG), Winter Barley (WBA), Winter Cereals (WCR), Winter Wheat (WWH) and Other (OTH). For spring and summer crops, the early surveys included the bare soil phase. When surveyors were able to link the bare soil structure to a sown crop class (e.g. ridges for POT, seedbed with characteristic row spacing for SBT or CAR) they were marked as stage 0 for such crop types. For non-specific bare soil stages (e.g. a ploughed or cultivated parcel), the generic Bare Soil (BSO) class was marked with the type of cultivation practice. Similarly, post-harvest observations may include again BSO stages or GMA if the parcel was re-sown with a catch crop. The sampling points were chosen on basis of geographical spread in the survey area and the possibility to stop safely. At every sam-
Fig. 2: The roof camera set-up is composed of two wide-angle side-looking cameras programmed to acquire one picture per second. A high definition camera is used to collect on the spot observations at 220 parcel locations along the surveyed track. The survey routing is supported with actualized Sentinel-2 false-colour imagery that serve as a map base on GPS-enabled mobile phone or tablet.

pling point, close-up and overview pictures were taken with the Nikon camera, while simultaneously identifying crop type and phenology stage. Phenological stages were recorded in accordance with the BBCH Monograph Stages (Meier, 1997) with some notable additions for BSO as explained above (introducing classes and generalizations described in section 2.3.2 and description of different stages in Supplementary Table S2).

2.2.3. Linking SLI to parcels

After the surveys, each picture and in-situ observation collected is linked to a specific parcel. Agricultural parcel geometries are openly available in the Netherlands as the 'Basisregistratie Gewaspercelen' (BRP) datasets (Nationaal Georegister - BRP, 2018) and were downloaded for the study. The in-situ observed data (section 2.2.2) were processed and attached to the corresponding parcel using a four-step geospatial process (Fig. 3).

1. The SLI captured locations (i.e. the GPS measurements on the road from which the pictures were taken) are first transposed 30 meters perpendicularly to the driving direction to the left or to the right side according to the camera (step 1). The geometrical intersection of the transposed locations with the agricultural parcel geometry allows linking most of the pictures to their corresponding parcel.

2. Similarly, the in-situ observations are transposed and linked to their corresponding parcels (step 2).

3. The parcels with in-situ observation were then selected and duplicate SLI at stop locations are removed (step 3).

4. Finally, the in-situ observation is appended to SLI on the basis of the date of the difference in time between when the picture acquisition and the in-situ observations (step 4).

2.3. Deep Learning

A deep learning framework was designed and implemented to classify the pictures in a given crop type and phenology class. In this section we describe 1) the choice of network, 2) the selection of training data, and 3) the training and hypertuning of the model.

2.3.1. Network selection

Instead of designing a deep learning network from scratch, we decided to use an out-of-the-box model and retrain only the final layer via transfer learning. Transfer learning is the process of taking a model trained for one task, where data is more readily available, and applying it to a new but similar task (Pan & Yang, 2009). The study makes use of a network, trained on the
large ImageNet data set (Deng et al., 2009), while bootstrapping a new fully-connected top layer to recognize other classes of images. In this sense the approach can be categorized as a network-based, multi-class deep transfer learning for feature extraction (Li & Hoiem (2017)), applied over an imbalanced test set (i.e. the classes have different number of pictures). Other than reducing data dependence, reducing time to develop the network, and the computational cost of training, this technique has been shown to advance performance in terms of a higher start, higher slope and a higher asymptote (Olivas et al., 2009).

The selected CNN comes from the MobileNet (Howard et al., 2017) family - a set of open-source, on-device, mobile-first computer vision models for TensorFlow. MobileNets are revolutionary in terms of the simplicity and efficiency of the network, achieving high accuracy results, while keeping the total number of parameters low. MobileNets use depthwise separable convolutions (Howard et al., 2017) instead of regular convolution reducing the number of parameters compared to the regular network with the same depth in the nets. This results in lightweight deep neural networks. The MobileNet version 2 used here represents an adequate tradeoff between accuracy and computing requirements as they include the residual connections and the expand/projection layers in their architecture (Sandler et al., 2018).

2.3.2. Training set selection
Following preliminary tests, we selected 400 images per class for training. To limit spatial auto-correlation in the sample, each parcel was sampled by a single image iteratively until the number of 400 images per class was reached. This process ensures that no parcel gets over-sampled due to its relatively large size (and consequently a larger number of images), compared to the other parcels of the same class. To create the training data set three steps were followed, 1) classes not separable visually were combined, 2) pictures taken at the edge of parcels were removed as they could contain two different crops and 3) pictures were visually inspected and re-labelled as ‘other’ when the crops was not visible.

Classes originally distinguished in the field were combined to avoid confusion as described in Table 1. Firstly, in the pre-emergence stage (i.e. BBCH stage is 0) of some crops, it is impossible to observe the canopy elements that would provide the basis for class distinction (however, sometimes the soil structure may contain clues); secondly when the difference between stages is minimal and not observable on SLI, and finally where the ripening and senescence phases are too similar to be distinguished. Importantly, these aggregations were done after inferring with the trained model multiple times, examining the confusion matrices, and applying common sense as to the level of detail that is discernible on a re-sized full resolution picture (see sec. 4.2).

A specific metric used to identify pictures close to the edge of the parcel was calculated on the basis of the distance between each SLI and the centroid of the corresponding parcel. This results in a ratio ranging between one (picture is further from the centroid) to $\frac{1}{2}$ (the picture is closest to the centroid). For most of the parcels, this ratio could be used to mark pictures potentially covering two parcels. When selecting images for the training set it is important to exclude those pictures, because they do not provide a good view on the crop. A value of 0.95 proved satisfactory.

In the third step pictures were visually inspected and re-labelled as ‘other’ when the crops was not visible. The ‘other’ class includes all the images that have a noticeable visual obstruction between the camera and the parcel. These can be in the form of dense vegetation, embankments, automobiles, trucks, or various stationary objects that are placed adjacent to
the road that prevent a direct view on the parcel. The 'other' class was manually selected and is included in the classification task, meaning it is one of the classes the network is trained on. Although there are many images with such visual obstructions present in the inference set, the study uses the 'other' class as a means of detecting, flagging, and removing such images before the classification task.

2.3.3. Training and hyper-tuning

The MobileNet architecture divides the training set into training, test, and validation bins. As such our original training sample benchmark of 400 images per crop phenological class is divided to 80% for training, 10% for testing, and 10% for validation. A hyper-tuning framework was implemented to find the relevant combinations of hyper-parameters. A schematic representation of the deep learning processing framework can be found in Fig. 4. Hyper-tuning a model is the process of systematically trying various parameter settings in order to find the best fit for the classification task.

Hyper-tuning was performed with the following parameters: type of MobileNet, number of training samples, batch size, optimizer, learning rate, learning momentum, and a series of image augmentations. Preliminary bench-marking indicated that the best performance, relative to training time was achieved with the primary model (width multiplier 1, and image size re-sampled to 224x224). For batch size, the best working values were batch sizes of 512 and 1024 images. For optimizers the experiment was ran using Adam and Gradient Descent (GD). A two-step process was applied for learning rate and momentum. First, initiate a random search within a 2D space with a good spread of values within the space. Second, models were trained in each combination to see in which domain of values the model performed best. The hyper-tuning was implemented through a bash script that simultaneously creates the directory tree, generates the files needed to run the network, and performs the relevant combinations of hyper-parameters.

2.4. Independent inferencing and parcel level classification

A total of 160 models were trained in parallel to identify the best combination of parameter values. After this first round of training, the top-three model configurations are identified based on the Macro F1 Score (M-F1) achieved during the independent inferencing. Then, a second round of training takes place on these three models including image augmentations to boost the accuracy of the best models. Finally, the best model is selected based again on a M-F1 ranking. A schematic representation of the processing steps for calculating accuracy is presented in Fig. 5.

A stand-alone inferencing with the best model was also performed on a large set of SLI. This allows us to evaluate the model’s performance independently from the training, test, and validation cycle in a real-life configuration. The classification accuracy is also evaluated when using majority voting at parcel level. Since multiple SLIs are taken at each parcel, at each time, an increased confidence in classification accuracy would improve the operational robustness of the methodology to collect in-situ data for cropped parcels.

Prior to calculating performance metrics on any level (Sections 3.3, 3.4) there is a reduction of the inference set. As noted in section 2.3.2, the introduction of an ‘other’ class is done in order to flag images without crop. To be able to cross compare model performance, it is necessary to maintain the same number of images. Therefore all images labelled as ‘other’ by any of the models, are identified and subtracted from the inference set prior to any accuracy calculation. The same was done for the bare soil class (‘BSO’).

2.4.1. Performance metrics

The presented study can be categorized as a multi-class im-balanced classification task, where all the classes matter equally (Sun et al., 2009). Because standard performance metrics, such as overall accuracy and error, have been shown to be misleading in such cases (Branco et al., 2015), the main metric, by which a comparison of model configurations is carried out, is the M-F1. The M-F1 gets calculated from a confusion matrix using the confusionMatrix function from R’s caret package. It outputs information on, among others, the following variables: Precision, Recall, F1. Precision (UA) is a measure of how many of the predicted positive outputs are actually positive, while recall (PA or sensitivity) is a measure of how many of the actual positives the model captured by labelling it as positive are so, and F1 is a balanced metric between the two. Because F1 is a score given to each class, the paper makes use of an aggregation of every per-class value into a single number, M-F1. M-F1 is calculated by taking the mean of all F1 scores, rather than taking the mean of the precision and recall and then calculating M-F1, in accordance with the literature (Opitz & Burst (2019)).

2.4.2. Performance levels and parcel aggregation

Performance metrics for computer vision classification tasks are usually calculated on a per-image basis. This means that for each image the network outputs a vector of probabilities with the likelihood for each class. This vector must sum to 1. For the 160 models, the M-F1 of each model is calculated by taking the mean of the F1 scores for the inference set for each class. Four performance levels are considered in this study. At the picture level (i.e. considering each image from the inference set), the M-F1 is computed for the phenological stage \((M - F1_{picture-\text{BBCH}})\) and for the crop type level by aggregating the pictures of the same crop type in the same class \((M - F1_{picture-crop})\). In addition to these two metrics, similar metrics are computed with an aggregation at the parcel level using a majority voting approach resulting in a M-F1 metric at the parcel for both phenological stage \((M - F1_{\text{parcelBBCH}})\) and crop type \((M - F1_{\text{parcel-crop}})\).

2.5. Code and Computational infrastructure

All the code developed for this study is available openly on the following repository: https://github.com/Momut1/
### Table 1: Class generalizations.

| New class | Old class | Description and rationale |
|-----------|-----------|---------------------------|
| BSO0      | BSO1, BSO3, BSO4, BSO5, BSO6, ONI0, POT1, SBT1, WWH0 | Although SLI has a potential to separate specific bare soil structures, e.g. by cultivation type, our current study focuses on vegetative crop stages. |
| GMA2      | BSO2      | Late season bare soil stage 2 is erroneously described by surveyors as Green manure. |
| GRA1      | GRA2, GRA3| One class for all post-emergence grass BBCH stages due to visual similarity. |
| ONI4      | ONI41     | Essentially the same class with minor difference, visible only upon very close-up inspection. |
| ONI48     | ONI45     | From 50% to 80% of expected bulb/shaft diameter visible. Detail not visible on such images. |
| SBT1      | SBT11     | Essentially the same class with minor difference, visible only upon very close-up inspection. |
| WWH8      | WWH9      | Ripening and senescence phases for winter wheat look very similar. Detail not visible on such images. |

---

**Fig. 4:** Deep learning processing framework. *Hyper-tuning* lists the parameters that the network is optimizing, *Training* relates to the training set selection (‘+1’ explained in Section 2.3.2) and training of the models, *Inference* shows the unbalanced inference set (explained in Section 2.4), and *Training with Augm* lists the augmentations applied for the second round of training over the best performing models, where they are selected based on the best from each performance level (see Section 2.4), for *Post-proc & Results* see Fig. 5.

**flevoland.** The working environment was set up as a docker image. The processing pipeline is fully automated for ease of reproducibility. It is fitted to work both locally and in a cloud through the use of a docker environment. Nearly the entire pipeline is made to function through the calling of dedicated shell scripts that separately accomplish the hyper-tuning, inferencing, and results calculation. For more information consult the *readMe* of the git repository. Part of the processing was done on the JRC “Big Data Analytics Platform” (BDAP), an in-house, cloud-based, versatile, petabyte-scale platform for heavy-duty processing (Soille et al., 2018). The offered GPU services work on a NVidia GeForce GTX 1080 Ti with 11GB memory, CUDA version 10.1, and CUDA driver version 418.67. Pre-processing, launching, and post-processing are done in the JEO-lab layer of the platform in a jupyter notebook docker container, running Tensorflow 1.3.0. Another part of the processing was done on Amazon Web Services (AWS) (Cloud, 2011) using the EC2 functionality to launch custom virtual machines. The study made use of one virtual machine for training which had the following specifications: G4dn EC2 instance with a NVIDIA T4 GPUs with 32GB memory with an identical CUDA and Cudnn version setting; and one for inferencing, which in turn had the following specifications: Inf1 EC2 instance with 24 CPUs.

### 3. Results

The results are divided into four sections. First, the resulting dataset, along with relevant processing, is presented (Section 3.1). Second, the best performing models are presented (Section 3.2). Thirdly, the ability of the model to discriminate between the respective crop and phenology combination (Section 3.3) will be shown. Results (M-F1) are presented both at image level and parcel level. Finally, the model ability to distinguish crop types only by aggregating the phenological stages of each crop into a single class is presented in Section 3.4.
3.1. Data collection

The collected data is organized in a structured form in a spatial database (PostgreSQL/Postgis) for easy querying and access. Each row represents an image, which is identifiable via a unique file path, also holding a variety of other attributes such as spatial location, capture timestamp, unique parcel ID, crop type according and the BBCH stage from the in-situ observation. A breakdown of image collection for the eight field campaigns can be seen in Fig. 6. The total number of collected pictures and the pictures of cropped parcels respectively amount to nine and five-fold the pictures with in-situ observations. Photos with in-situ observations (green bar) amount to an overall total of 48,444 and are the result of the geospatial processing described in Section 2.2.3.

An example of random pictures taken for each of the 17 different crops is shown in Fig. 7. These examples of pictures show that the crop parcel of interest covers only a small fraction of the whole scene. Indeed, mainly the sky and the road along with other smaller elements such landscapes features (e.g. trees, hedges, buffer strips) or artificial elements (e.g. building, windmill) could cover a large part. These pictures also illustrate the wide angle lens distortion effect from the cameras allowing to capture an almost 180 degree azimuthal view.

The number of images and parcels per crop and per BBCH stage that make up the complete labelled dataset are shown in Fig. 8. The highlighted bars are the classes that make up the computer-vision experiment while the grey ones were discarded as not reaching the 400 pictures required thresholds. The vertical bars represent the number of pictures for each BBCH stage of each crop, the connected points represent the number of parcels (on the secondary Y-axis) for each BBCH stage of each crop. After discarding the classes with less than 400 pictures and applying the class generalizations from Table 1, we obtain a final training dataset of 21 crop/BBCH combination classes which aggregate at crop level to 12 crop classes (including ‘other’). Specific classes (i.e. ONI4, SBT1, SCR3, MAI1) with more than 400 images were removed due to distance to centroid and visual screening during training set selection.

The picture dataset is available at https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/DRLL/FlevoVision/training/.

3.2. Best model selection

The best model configuration was selected based on the M-F1 ranking (detailed results can also be found in GitHub here). Table 2 summarizes the top three performing models; the most accurate is # 44 with a M-F1 of 59.4%, 86.9%, 62.3%, 88.1% respectively for the picture-phenology, parcel-phenology, picture-crop, and parcel-crop (Section 2.4.2). The best model configuration (Model # 44) was built with a GD optimizer, a training batch size of 512, a learning rate of 0.08990006381198408, with augmentations flipping images left and right, training up to the last to final 2999th epoch. The model is also the best in not confusing bare soil with the other classes. This confusion occurs less than 3.2 times compared to both other top models.

3.3. Phenology-level recognition

The best model (#44) is applied on the independent inference set (12,140 pictures on 482 parcels). The confusion matrices are displayed in Fig. 9 for both picture (Fig. 9a) and parcel (Fig. 9b) level. As described in section 2.4, the bare soil (‘BSO0’) and other (‘OTH0’) classes are removed to assess the performance resulting in a 19-class confusion matrix. At the picture level, the model reached a M-F1 of 59.4%, while 86.9% was achieved when aggregating to parcel-level, an increase of
27.5%. At the image level, the model is certainly more likely to make more errors of commission than of omission, which is reflected by the relatively lower values of UA compared to those of PA (see Fig. 9b). While PA is, for the majority of classes, above 70% (excluding GMA2, GRA1, ONI48, POT6, POT8, SBT39, WWH3), the UA score exceeds 70% for only 4 classes (CAR1, CAR4, GRA1, SBT39). Especially poor in this case are the cereal classes, spring cereal (SCR2) and WWH, exhibiting confusion between themselves, and with GRA. This is also reflected in the low UA for these classes, meaning the model was miss-labelling many images from other classes as belonging to either of the cereal classes. Furthermore the distinction between cereals is difficult before the crop has elon-gated to form more recognisable canopy elements (e.g. ears). SCR is actually a composite class used in the early surveys. Most parcels in these class evolve into distinct SBA and SWH in later surveys. Indeed it is the case that the class with the lowest PA, GRA1, is the one that is most often confused with other classes. At the parcel level, however, the confusion is significantly less with PA values in the high 80s and 90s. Interestingly, the UA trend does not follow the same pattern at parcel level. For example, the class with the second to worst UA at the picture level (WWH7 with 29.3), has the best UA on the parcel level (100).

### 3.4. Crop-level recognition

A total of 10 crops occur in the 19 crop-BBCH combinations. Performance at crop level is evaluated by running the model used to detect the phenological stages, and subsequently keeping the classification of the most likely crop class removing confusion in between the BBCH stages of the same crop. Model configuration 44 yielded a M-F1 of 62.3% on the picture level and 88.1% on the parcel level, meaning an increase of 25.8%. At picture level, there are both errors of commission and omission, reflected by the relatively low values of PA and UA. As with the parcel level performance on BBCH, the error decreases significantly from picture to parcel. The major part of the confusion seems to stem from the grass (GRA), green

---

**Table 2: Output for the top three performing models with augmentations.** The applied augmentation was a left-right flip. The table shows (in order) the model number, along with the relevant configuration (Learning rate, Batch size, Momentum, Optimizer), the number of images classified as other and bare soil for each model, the training and validation accuracy, and the M-F1 for each aggregation level.

| Ranking | 1     | 2     | 3     |
|---------|-------|-------|-------|
| Model # | 44    | 107   | 127   |
| Learning Rate | 8.99e-02 | 4.6e-04 | 6.58e-04 |
| Batch Size | 512   | 1024  | 1024  |
| Momentum | 0e+00 | 9.36e-01 | 9.19e-01 |
| Optimizer | GD    | Adam  | Adam  |
| Other # | 668   | 663   | 661   |
| Bare soil # | 867   | 2166  | 2116  |
| Validation Accuracy | 81.3  | 78.5  | 78.5  |
| Training Accuracy | 96.7  | 93.0  | 93.0  |
| $M - F_1^{\text{picture-BBCH}}$ | 59.4  | 57.4  | 57.8  |
| $M - F_1^{\text{parcel-BBCH}}$ | 86.9  | 86.0  | 85.6  |
| $M - F_1^{\text{picture-CROP}}$ | 62.3  | 60.4  | 60.9  |
| $M - F_1^{\text{parcel-CROP}}$ | 88.1  | 88.2  | 87.5  |

---

**Fig. 6: Number of image collected of the eight field campaigns from March 2018 to October 2018.**
manure (GMA), and winter wheat (WWH) classes, whereby parcels from other crops get classified as either grass or winter wheat thus dropping the UA of these classes; and green manure parcels get classified themselves as belonging to other classes, thereby lowering the PA of this class.

4. Discussion

4.1. Context

Similar studies have recently been carried out although direct comparison is difficult as protocols and crops surveyed vary significantly. Wu et al. (2021) obtained an overall accuracy of 91.1% for 12 crop types (bare land, cotton, maize, peanut, rape-seed, rice, sorghum, soybean, sunflower, tobacco, vegetable, wheat) using majority voting for five classifier benchmarks. However, overall accuracy metrics are probably not the best for imbalanced datasets (Branco et al., 2015). A study by Yan & Ryu (2021) explored Google Street View with deep learning for crop type mapping resulting in an accuracy ranging from 92% for 8 crops (alfalfa, almonds, corn, cotton, grape, other, piscatichios, rice) to 97% for 3 crops (corn, other, soybean). While the crops and metrics used are very different, the measured performance of Wu et al. (2021) and Yan & Ryu (2021) are higher. Further studies and open datasets for benchmarking different methods are needed to cross-compare the results more consistently.

In any case, the study is a push towards a new state of the art in combining existing approaches into a methodology that can be easily scaled and executed for a reasonable cost. While other work has previously made contributions to plant recognition (Ringland et al., 2019) and phenology recognition (Yalcin, 2015), and devising smart massive image-capture for computer vision (Wu et al., 2021), there has not been a push towards having phenology level information collected with such methods. In terms of technical novelty, the work shows an automated method for a much easier and faster massive image labeling (section 2.2.3). Additionally, the process for performing parcel-level classification through majority voting, multiple modes, and maximum probability summing, albeit simplistic,
Fig. 8: Sample of street-level images for which in-situ information was derived from field observations. The vertical bars represent the number of pictures for each BBCH stage of each crop (see Supplementary Table S2 for label details), the connected points represent the number of parcels (on the secondary Y-axis) for each BBCH stage of each crop, and the dotted red line represents the threshold of 400 images. The highlighted bars are the classes that make up the computer-vision experiment. The bars themselves are also stacked to reflect the class generalizations, described in Table 1.
Fig. 9: Confusion matrices for phenology recognition (19 BBCH classes) at picture and parcel level. PA and UA are in the margins. See Supplementary Table S2 for label details.
provides near-real-time parcel relevant information. This could be of major use to single or organizations of farmers, to agri-orientated consultancies, or to regulatory bodies.

4.2. Limitations

Although novel in theoretical, semantic, and technical regards, the study nevertheless suffers from some notable shortcomings and drawbacks.

Firstly, in the proposed approach, the input data loses the majority of its detail due to the network’s inability to handle full resolution images. It is well known that current deep learning architectures use input images that are of low resolution. For instance in radiography, the optimal input image size is between 256 and 448 pixels per dimension (Sabottke & Spieler (2020)), and in general the trend is followed in the agricultural domain. The MobileNet V2 (Sandler et al., 2018) used here automatically re-scales all input images to the largest resolution it can handle, which is 224 pixels per dimension. The full resolution images collected for this study are of dimensions 3840 pixels in width to 2160 pixels in height. Thus each image gets re-scaled from its parallelogram shape 17.1 times in width and 9.6 times in height to produce the rectangular input image that the network actually trains on. With such a huge reduction and loss of information, it is not surprising that much confusion exists between the relevant classes (see Section 4.4 for recommendations for future studies). Nonetheless, models that correctly distinguished ‘bare soil’ from other classes were penalized. Further developments could consider hierarchically classifying bare soil and crop in a first step, and secondly distinguishing different types of bare soil e.g. related to ploughing patterns associated to crops.

Secondly, the way the workflow deals with the ‘other’ and ‘bare soil’ classes can be improved (Section 2.4). A total of 3,065 images were excluded consisting of 2,361 labeled ‘bare soil’ and 704 labeled ‘other’ respectively to facilitate accuracy calculations across models. Excluding ‘bare soil’ images increased the performance of the vast majority of model configurations. For each performance level (\(M - F_1^{picture-BBCH}\), \(M - F_1^{parcel-BBCH}\), \(M - F_1^{picture-crop}\), and \(M - F_1^{parcel-crop}\)), the increase was respectively 1.9%, 3.6%, 2.3%, and 5.5%. This suggests that ‘bare soil’ is either significantly more present in the dataset and is ergo a disproportionately imbalanced class, or that it is too hard to be distinguished from seedlings. Nonetheless, models that correctly distinguished ‘bare soil’ from other classes were penalized. Further developments could consider hierarchically classifying bare soil and crop in a first step, and secondly distinguishing different types of bare soil e.g. related to ploughing patterns associated to crops.

Thirdly, the study assumes a uniform distribution of a BBCH stage across the entire roadside perimeter of the parcel. In reality, the detailed BBCH observation was made by the surveyor at a particular point in the parcel. Due to the nature of the logic described in section 2.2.3, the processing chain assumes that all images associated with the parcel exhibit the same stage. In doing so, we may be introducing some noise by grouping together pictures that actually show the crop in a variety of stages.

4.3. Parcel-level interpretation

Estimating parcel-level model performance is a key aspect of this study. Specifically, the questions are what is the minimum
Pictures Per Parcel (PPP) needed for a correct classification, and after how PPP is a correct classification certain?

Although the number of surveyed parcels is 307 (section 2.1), the parcels over which results are calculated on BBCH and crop levels are 482 (Fig. 9b) and 326 (Fig. 10b) respectively. Rather than aggregating on the unique combination of crop type and survey date, the study chose to perform the aggregation based on crop type and parcel ID. The latter combination is not unique and thus parcel IDs get re-coded to reflect the different BBCH stages (for single crop parcels) and crop types (for multi-crop parcels) of the parcel. This method is preferable, as it deals simultaneously with cases of more prevalent BBCH stages (SBT39) and sparsely represented classes.

The histogram on top of Fig. 11 shows the frequency of correctly and incorrectly classified parcels in terms of the number of PPP. The data is skewed to the right, meaning that the majority of incorrect and correct classifications occur at the lower ranges. The mean of incorrectly and correctly classified parcels is 9.6 and 27.4 PPP respectively. It is thus impossible to answer the first question, as there are both correctly and incorrectly classified parcels with a single picture. Alternatively, the data gives insight into the probability for a correct classification of a parcel with a given PPP. This information is shown on the connected scatter plot on the right of Fig. 11. The histogram also illustrates the evolution of certainty in getting a correct classification with regards to natural breaks. In the range of 0-25 PPP, the proportion of correct to incorrect parcel classifications are 83.4% and 16.6% respectively; in the range of 25-50 PPP the percentages are 96% and 4%; finally for parcels with over 50 pictures the classifications are 100% correct. The figure in the middle gives a more detailed breakdown of the other two. Firstly for almost all the classes, the range of values for correctly classified parcels is larger than for the incorrect ones, excluding POT6. However, for six out of nineteen classes (MAI7, ONI48, POT6, SBT14, WWH2, WWH3) the mean number of PPP is higher for the incorrectly classified parcels. With such a spread between the anomalous classes it is impossible to attribute this to a specific crop or set thereof. From visually inspecting the locations of the miss-classified parcels, there is no spatial clustering. This suggests that, especially in the lower ranges of PPP (below 25), image abundance cannot solely guarantee correct classifications, but rather that there is also a defining role for image quality.

To gain insight into the role of image quality, the pictures of all 59 miss-classified parcels at BBCH level were visually inspected. The errors are due to environmental, geographic, or processing related factors. The first category includes issues like sun glare (Fig. 12a), and tall grass or linear elements obstructing the view on the parcel (Fig. 12b). The second category includes situations where the camera is away from the parcel (Fig. 12c), and where parcels are imaged from different aspects. In the third category we find issues related to the distance to centroid ratio filter not performing adequately (Fig. 12d), resulting in many pictures of parcel boundaries being included. In this class errors related to labelling are also included, an example is shown in Fig. 12e where parcel 2295159 is labelled as SBT39, but classified as ONI48. Comparing this example with the SBT39, as shown in Fig. 7, a clear difference is visible. Finally, we include here pictures that are blurred or out of focus (Fig. 12f). For parcels that are incorrectly classified, more than one factor could have led to an incorrect classification. A parcel may be both obscured by tall grass and also far away from the camera. Another point which comes up often is the relatively few pictures per parcel that are available. Although most of the miss-classified parcels exhibit one or more of these drawbacks, it must be noted that for a significant number of them (20 out of 59) there is nothing visibly wrong with the pictures and hence the incorrect classification is purely due to sub-optimal model performance.

4.4. Recommendations

The availability of frequent and high resolution Copernicus Sentinel satellite data streams requires complementary in-situ data for instantaneous cross-referencing and validation. To boost collection efficiency, novel ways to collect or derive such high quality reference data need to be explored. In-situ data collection is a tedious and expensive task and is often limited to small scales, and specific purposes, serving various scientific validation experiments. Only very limited amounts of standardized reference data end up in the open data domain, notwithstanding collaborative community efforts like the Joint Experiment of Crop Assessment and Monitoring (JECAM) (Kussul et al., 2014)) of the Group on Earth Observation (GEO). Efforts to use computer vision approaches on the archive of large scale standardized in-situ statistical surveys including geo-tagged pictures could provide new information on historical dynamics of land cover and land use. A recent harmonisation of the LUCAS (d’Andrimont et al., 2020b) survey has delivered 5.4 millions of land-cover-labeled pictures over 15 years statistically distributed in EU. More research is needed to apply state-of-the-art computer vision methods on such datasets.

For the current study, tiling the full resolution images into 224x224 pure land cover tiles should significantly improve the model’s performance. The challenge is to detect the parcel of interest in the landscape SLI. This could be done by detecting the horizon line and then taking a buffer in a down-ward direction from the horizon to a relevant threshold that would encompass the pixels exhibiting only the parcel. This buffered group of pixels would then be split into 224x224 tiles in order to create better input images for the network. Alternatively, specific semantic segmentation nets such as Detectron2 (Wu et al., 2019) could be trained to detect the parcel and tiles could be extracted in a similar manner from the parcel segmentation mask. Such tiling would remove the problem of mixed classes inherent to the SLI and would increase the number of pictures available as many tiles could be extracted from one SLI.

Last but not least, recent methodological developments for botanical taxonomic inventory using massive amount of pictures collected by crowdsourcing and experts such as PlantNet (Affouard et al., 2017) could provide method, tools (i.e. API) taking benefit of a wide community. These methods and tools should be investigated for classification of SLI.
5. Conclusions

In this paper, we have presented a framework to collect and extract crop type and phenological information from street level imagery using computer vision. We have shown how, during the 2018 growing season, high definition pictures were captured with side-looking action cameras in the Flevoland province of the Netherlands. Thanks to a classification performed with transfer learning of convolutional neural network modules along with a hypertuning methodology, we discriminate crop type with a Macro F1 score of 88.1% and main phenological stage at 86.9% at the parcel level. While a lot of developments are still needed to improve the performance of the classification, the proposed framework is a first step in speeding up high quality in-situ data collection and opens avenues for massive data collection via automated classification using computer vision.

6. Author contributions

R.D, M.Y., and M. V. conceptualized the study and designed the methodology. R.D., M.Y., L.M. processed the data. R.D., M.Y., L.M., M. V. analyzed the data and wrote the paper.

7. Acknowledgements

The authors thank Guido Lemoine for his support and agro-nomic inputs. The authors would like to thank TerraSphere Imaging for the high quality data collection, in particular Menno de Vries and Paul van der Voet. The authors would also like to thank Dominique Fasbender for his precious statistical advices and the JRC colleagues from the Big Data Analytics project for their support. The author also thank the Exploratory Research program of the Joint Research Centre for their support within the Rural Refocus project (31280).

8. References

References

Affouard, A., Goèau, H., Bonnet, P., Lombardo, J.-C., & Joly, A. (2017). Plantnet app in the era of deep learning. In ICLR: International Conference on Learning Representations.

Anami, B. S., Malvade, N. N., & Palaiah, S. (2020). Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images. Artificial Intelligence in Agriculture, 4, 12–20.

Barve, V. (2014). Discovering and developing primary biodiversity data from social networking sites: A novel approach. Ecological Informatics, 24, 194–199.

Branco, P., Torge, L., & Ribeiro, R. (2015). A survey of predictive modelling under imbalanced distributions. arXiv preprint arXiv:1505.01658.

Cao, M., Sun, Y., Jiang, X., Li, Z., & Xin, Q. (2021). Identifying leaf phenology of deciduous broadleaf forests from phenocam images using a convolutional neural network regression method. Remote Sensing, 13, 2331.

Champ, J., Mora-Fallas, A., Goèau, H., Mata-Montero, E., Bonnet, P., & Joly, A. (2020). Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots. Applications in plant sciences, 8, e11373.

Cloud, A. E. C. (2011). Amazon web services. Retrieved November 9, 2011.

d’Andrimont, R., Lemoine, G., & Van der Velde, M. (2018). Targeted grassland monitoring at parcel level using sentinels, street-level images and field observations. Remote Sensing, 10, 1300.

d’Andrimont, R., Taymans, M., Lemoine, G., Ceglar, A., Yordanov, M., & van der Velde, M. (2020a). Detecting flowering phenology in oil seed rape parcels with sentinel-1 and-2 time series. Remote sensing of environment, 239, 111660.

d’Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eisselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Jöbges, C., Lemoine, G. et al. (2020b). Harmonised lucas in-situ land cover and use database for field surveys from 2006 to 2018 in the european union. Scientific Data, 7, 1–15.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248–255). Ieee.
Fig. 12: Illustrations from various miss-classified parcels of model 44 showcasing the different types of shortcomings that can exist, which make for an incorrect parcel classification.
architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8697–8710).
Contents

1 Introduction ................................................................. 1
  1.1 Objectives ............................................................. 2

2 Materials and Methods ..................................................... 2
  2.1 Study area ............................................................. 2
  2.2 Field Data Collection .................................................. 2
    2.2.1 Street-level Imagery acquisition ................................ 2
    2.2.2 Phenology Observations ........................................... 3
    2.2.3 Linking SLI to parcels ............................................. 4
  2.3 Deep Learning .......................................................... 4
    2.3.1 Network selection ................................................... 4
    2.3.2 Training set selection .............................................. 5
    2.3.3 Training and hyper-tuning ......................................... 6
  2.4 Independent inferencing and parcel level classification .......... 6
    2.4.1 Performance metrics ............................................... 6
    2.4.2 Performance levels and parcel aggregation ....................... 6
  2.5 Code and Computational infrastructure ................................ 6

3 Results ........................................................................... 7
  3.1 Data collection .......................................................... 8
  3.2 Best model selection ..................................................... 8
  3.3 Phenology-level recognition ............................................ 8
  3.4 Crop-level recognition .................................................. 9

4 Discussion .................................................................. 10
  4.1 Context .................................................................. 10
  4.2 Limitations ............................................................. 13
  4.3 Parcel-level interpretation ............................................. 13
  4.4 Recommendations ..................................................... 14

5 Conclusions ................................................................. 15

6 Author contributions ...................................................... 15

7 Acknowledgements ........................................................ 15

8 References ................................................................. 15

List of Figures

Fig. 1 The study area is in the province of Flevoland in the Netherlands. The specific study area includes the Oost Flevoland and Noord-Oost Polder districts. During 8 surveys a total of 1,691 in-situ crop phenology observations were made at 220 parcel locations alongside the 200-km surveyed track. Projection of the map: Amersfoort / RD NEW (EPSG:28992). ................................................................. 3

Fig. 2 The roof camera set-up is composed of two wide-angle side-looking cameras programmed to acquire one picture per second. A high definition camera is used to collect on the spot observations at 220 parcel locations along the surveyed track. The survey routing is supported with actualized Sentinel-2 false-colour imagery that serve as a map base on GPS-enabled mobile phone or tablet. ................................................................. 4

Fig. 3 Four-step spatial analysis to link observation data and SLI located on the road to specific agricultural parcels. ............................................................ 4

Fig. 4 Deep learning processing framework. Hyper-tuning lists the parameters that the network is optimizing, Training relates to the training set selection (‘+1’ explained in Section 2.3.2) and training of the models, Inference shows the unbalanced inference set (explained in Section 2.4), and Training with Augm lists the augmentations applied for the second round of training over the best performing models, where they are selected based on the best from each performance level (see Section 2.4), for Post-proc & Results see Fig. 5. ................................................................. 7
Fig. 5  The output predictions of the best model provide the probability for each picture to be in the different BBCH classes. Performance metrics are computed for both crop type and phenological stage and predictions are evaluated at picture and parcel level. Finally, the confusion matrices, PA, UA, F1 and the M-F1 results are computed for all performance levels. Resulting figure references are in blue.

Fig. 6  Number of image collected of the eight field campaigns from March 2018 to October 2018.

Fig. 7  Examples of image for each of the 17 different collected crop types for random BBCH stage. In addition to the crop field, the picture could included sky, road, buffer strips, trees, woody features, buildings or other elements of the landscape. The label of the crop along the BBCH is shown on the upper left of the image (see Supplementary Table S2 for label details).

Fig. 8  Sample of street-level images for which in-situ information was derived from field observations. The vertical bars represent the number of pictures for each BBCH stage of each crop (see Supplementary Table S2 for label details), the connected points represent the number of parcels (on the secondary Y-axis) for each BBCH stage of each crop, and the dotted red line represents the threshold of 400 images. The highlighted bars are the classes that make up the computer-vision experiment. The bars themselves are also stacked to reflect the class generalizations, described in Table 1.

Fig. 11 Distribution of correctly and incorrectly classified parcels. The three plots are derived from the same table and show a different aspects of the data. The top plot is a histogram of the number of pictures per parcel, the scatter plot to the right shows the probability of a parcel with a certain number of pictures to be classified correctly, and the barplot in the middle shows a breakdown per class of the data with the numbers over each bar indicating the number of parcels that belong to each group of correctly or incorrectly classified parcels.

List of Tables

Table 1  Class generalizations.

Table 2  Output for the top three performing models with augmentations. The applied augmentation was a left-right flip. The table shows (in order) the model number, along with the relevant configuration (Learning rate, Batch size, Momentum, Optimizer), the number of images classified as other and bare soil for each model, the training and validation accuracy, and the M-F1 for each aggregation level.
**Supplementary Material**

Supplementary Table S1: Action camera settings for SLI collection.

| Settings          | Used parameters                |
|-------------------|--------------------------------|
| Zoom              | ON – 1.1                       |
| Shooting          | Time-lapse                     |
| Time lapse shooting interval | 1 sec                       |
| Loop recording mode | Continue                     |
| Shooting scene settings | SCN N                        |
| Angle             | Wide                           |
| White balance     | Automatic                      |
| Color mode        | Natural                        |
| Automatic exposure | AE – T                        |

Supplementary Table S2: Table explaining the meaning behind all the present BBCH stages of all surveyed crops. The table can be read as a detailed view of Fig. 9.

| Crop BBCH | Crop common name | BBCH | Description |
|-----------|------------------|------|-------------|
| BSO0      | Bare soil        | 0    | Untreated bar |
| BSO1      | Bare soil        | 1    | Untreated stubble |
| BSO2      | Bare soil        | 2    | Green manure |
| BSO3      | Bare soil        | 3    | Ploughed/converted |
| BSO4      | Bare soil        | 4    | Harrowed |
| BSO5      | Bare soil        | 5    | Ploughed and harrowed |
| BSO6      | Bare soil        | 6    | Seedbed prepared code |
| CAR0      | Carrot           | 0    | Germination |
| CAR1      | Carrot           | 1    | Leaf development |
| CAR2      | Carrot           | 2    | No info |
| CAR3      | Carrot           | 3    | No info |
| CAR4      | Carrot           | 4    | Development of harvestable vegetative plant parts |
| GMA0      | Green manure     | 0    | No info |
| GMA1      | Green manure     | 1    | No info |
| GMA2      | Green manure     | 2    | No info |
| GMA3      | Green manure     | 3    | No info |
| GMA4      | Green manure     | 4    | No info |
| GRA0      | Grassland        | 0    | Bare soil |
| GRA1      | Grassland        | 1    | After mowing height 0-5cm |
| GRA2      | Grassland        | 2    | Leaves not yet bending, height 5-20cm |
| GRA3      | Grassland        | 3    | Before mowing, bending (height 20-30cm) |
| GRS7      | Grass seeds      | 7    | No info |
| MAI0      | Maize            | 0    | Germination |
| MAI1      | Maize            | 1    | Leaf development |
| MAI3      | Maize            | 3    | Stem elongation |
| MAI5      | Maize            | 5    | Inflorescence emergence, heading |
| MAI6      | Maize            | 6    | Flowering, anthesis |
| MAI7      | Maize            | 7    | Development of fruit |
| MAI8      | Maize            | 8    | Ripening code |
| ONI0      | Onion            | 0    | Germination |
| ONI1      | Onion            | 1    | Leaf development (main shoot) |
| ONI41     | Onion            | 41   | Leaf bases begin to thicken or extend |
| ONI45     | Onion            | 45   | 50% of the expected bulb or shaft diameter reached |
| ONI48     | Onion            | 48   | Leaves bent over in 50% of plants |
| ONI49     | Onion            | 49   | Leaves dead, bulb top dry; dormancy |
| POT0      | Potato           | 0    | Germination |
| POT1      | Potato           | 1    | Leaf development |
| POT2      | Potato           | 2    | Formation of basal side shoots below and above soil surface (main stem) |
| POT3      | Potato           | 3    | Main stem elongation (crop cover) |
| POT5      | Potato           | 5    | Inflorescence (cyme) emergence |
| POT6      | Potato           | 6    | Flowering |
| POT8      | Potato           | 8    | Ripening of fruit and seed |
| POT9      | Potato           | 9    | Senescence code |
| WCR0      | Cereal           | 0    | Germination |
| WCR1, SCR1| Cereal           | 1    | Leaf development |
| SCR2, WWH2| Cereal           | 2    | Tillering |
| SBA3, SCR3, WWH3 | Cereal | 3    | Stem elongation |
| SBA5, SWH5, WWH5 | Cereal | 5    | Inflorescence emergence, heading |

*Continued on next page*
| Crop BBCH | Crop common name | BBCH | Description                                      |
|-----------|-----------------|------|--------------------------------------------------|
| SBA7, SWH7, WWH7 | Cereal | 7    | Development of fruit                             |
| SBA8, SWH8, WWH8, WBA8 | Cereal | 8    | Ripening                                         |
| SBA9, SWH9, WWH9 | Cereal | 9    | Senescence code                                  |
| SBT0      | Sugar beet      | 0    | Germination                                      |
| SBT11     | Sugar beet      | 11   | Leaf development - first leaf unfolded           |
| SBT14     | Sugar beet      | 14   | Leaf development - 3 leaves unfolded             |
| SBT19     | Sugar beet      | 19   | Leaf development - 9 or more leaves unfolded     |
| SBT2      | Sugar beet      | 2    | No info                                          |
| SBT32     | Sugar beet      | 32   | Leaves cover 20% of ground                       |
| SBT39     | Sugar beet      | 39   | Crop cover complete: leaves cover 90% of ground code |
| TLP0      | Tulip           | 0    | No info                                          |
| TLP12     | Tulip           | 12   | No info                                          |
| TLP45     | Tulip           | 45   | No info                                          |
| TLP48     | Tulip           | 48   | No info                                          |
| TLP49     | Tulip           | 49   | No info                                          |
| TLP6      | Tulip           | 6    | No info                                          |
| VEG1      | Vegatable       | 1    | Leaf development (Main shoot)                    |
| VEG2      | Vegatable       | 2    | Formation of side shoots                         |
| VEG8      | Vegatable       | 8    | Ripening of fruit and seed                       |