NPTEL ONLINE CERTIFICATION COURSES

Machine Learning for Soil and Crop Management

Prof. Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology Kharagpur

Week 8: UAV AND ML APPLICATIONS IN AGRICULTURE

LECTURE 36

CONCEPTS COVERED

- IMAGE BASED SOIL PROPERTY PREDICTIONS
- CROP IMAGE+ CNN FOR WEED IDENTIFICATION

KEYWORDS

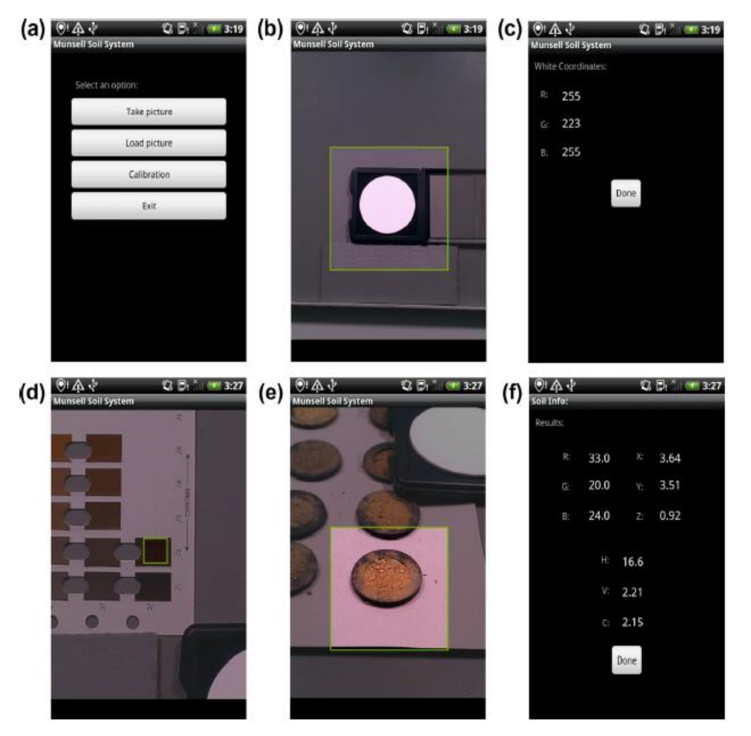
- Weed
- CNN
- VGC16
- Stride
- Padding

CAN SMARTPHONE REPLACE A MUNSELL SOIL COLOR CHART?

1. Custom image-processing app for smartphone

2. Model for RGB---Munsell notations

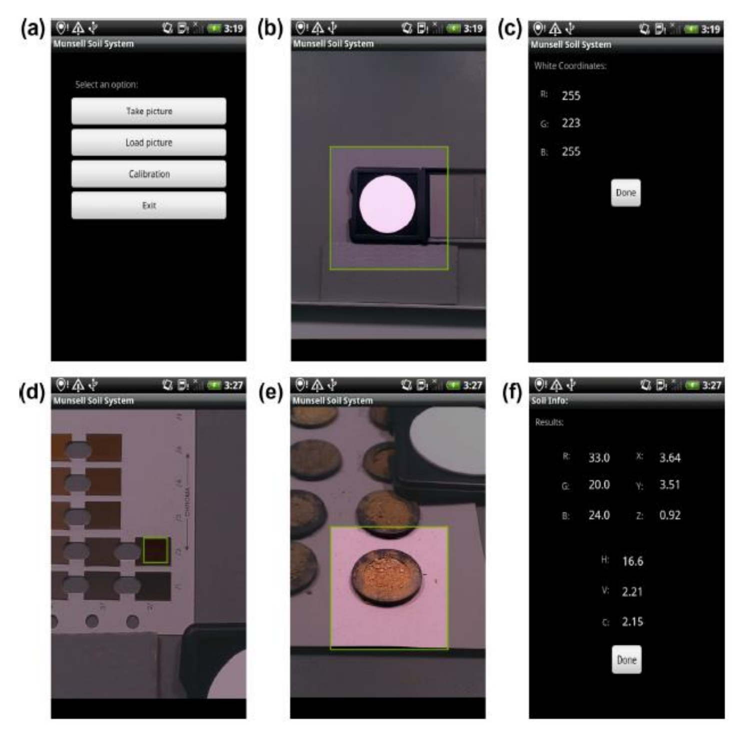
3. Mobile software: image processing + model RGB to Munsell



Gomez-Robledo et al. (2013)

CAN SMARTPHONE REPLACE A MUNSELL SOIL COLOR CHART?

Developed an Android application that takes a picture of a soil sample, allowing the user to select the region of interest and then, after a RGB image-processing and a polynomial process transform between colour spaces, the Munsell (HVC) and CIE (XYZ) coordinates appear on the screen of mobile phone.



Gomez-Robledo et al. (2013)

CROP IMAGES + CNN FOR WEED IDENTIFICATION

Table 1. Plant species used in the training of the CNNs, the relevant EPPO code, along with the numb
of labels per plant species in total and for the training, validation and training subset

Plant Species	EPPO CODE	Total Images	Train Images	Validation Images	Testing Images
Alopecurus myosuroides Huds.	ALOMY	7423	5196	1113	1114
Amaranthus retroflexus L.	AMARE	5274	3691	791	792
Avena fatua L.	AVEFA	12,409	8686	1861	1862
Chenopodium album L.	CHEAL	2690	1882	403	405
Helianthus annuus L.	HELAN	16,426	11,498	2463	2465
Lamium purpureum L.	LAMPU	7603	5322	1140	1141
Matricaria chamomila L.	MATCH	15,159	10,611	2273	2275
Setaria spp. L.	SETSS	2378	1664	355	359
Solanum nigrum L.	SOLNI	2979	2085	446	448
Solanum tuberosum L.	SOLTU	2742	1919	411	412
Stellaria media Vill.	STEME	6941	4858	1041	1042
Zea mays L.	ZEAMX	11,106	7774	1665	1667
	SUM	93,130	65,186	13,962	13,982

Sony Alpha 7R Mark4

Peteinatos et al. (2013)

CROP IMAGES + CNN FOR WEED IDENTIFICATION

- For each image, a binary image was created, using the Excess Green–Red Index as a thresholding mechanism to separate plant material from the soil

- Each connected pixel formation from this thresholding procedure consisted of a potential region of interest that should be fed in the CNNs and was separated and prelabeled, creating the relevant bounding box

Peteinatos et al. (2013)

CROP IMAGES + CNN FOR WEED IDENTIFICATION

- These labels were examined by a human expert who discarded possible wrong classifications or unwanted weeds
- Each connected pixel formation from this thresholding procedure consisted of a potential region of interest that should be fed in the CNNs and was separated and prelabeled, creating the relevant bounding box

Peteinatos et al. (2013)

CROP IMAGES + CNN FOR WEED IDENTIFICATION

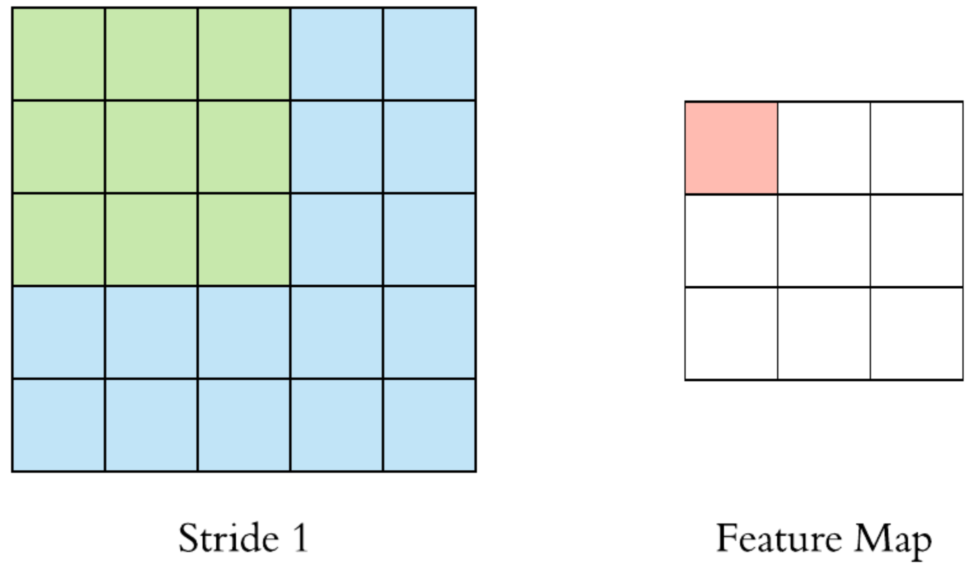
- VGC16
    - A CNN architecture
    - The best performing networks at the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition
    - Considered to be one of the excellent vision model architecture till date
    - Can provide a high performance and respective accuracies, even when the image datasets are small
    - Input: 3 channel RGB crop image
    - Total 16 layers: 13 convolutional and 3 FC layers
    - Less computer intensive than other networks

CROP IMAGES + CNN FOR WEED IDENTIFICATION

- VGC16 architecture
- Input: Image (224,224, 3)

CROP IMAGES + CNN FOR WEED IDENTIFICATION

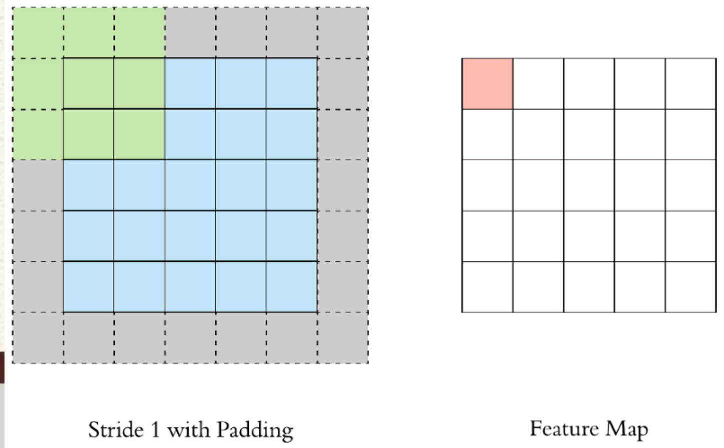
- **Stride**
- **Stride denotes how many steps we are moving in each steps in convolution**
- **Default value= 1**



Source: dshahid380 https://towardsdatascience.com/

CROP IMAGES + CNN FOR WEED IDENTIFICATION

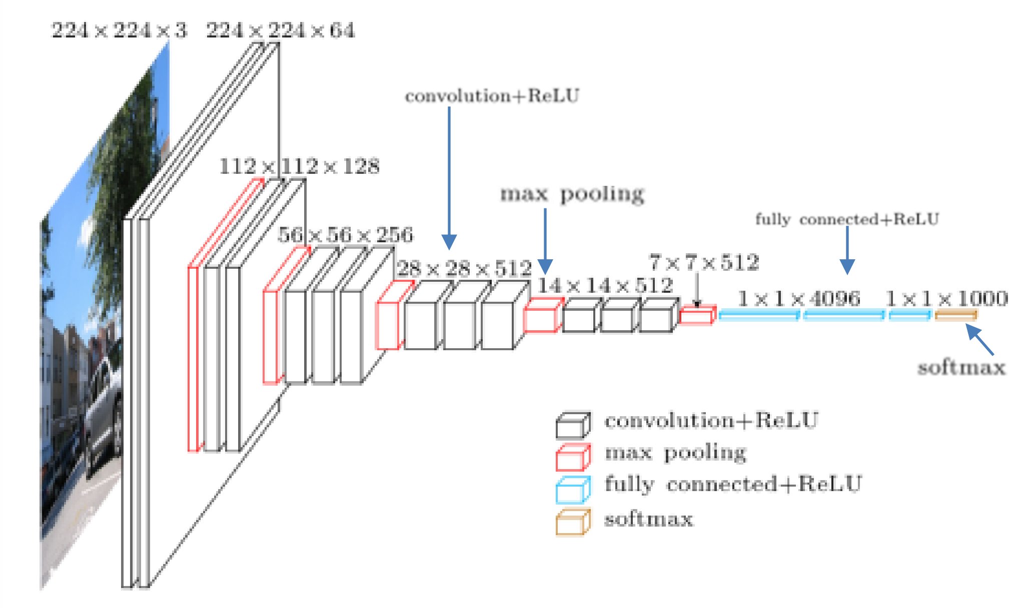
- Padding: Extends the area of an image in which a CNN processes
- The kernel/filter which moves across the image scans each pixel and converts the image into a smaller image
- In order to work the kernel with processing in the image, padding is added to the outer frame of the image to allow for more space for the filter to cover in the image
- SAME padding: with a stride of 1, the layer's outputs will have the same spatial dimensions as its inputs



Source: dshahid380 https://towardsdatascience.com/

CROP IMAGES + CNN FOR WEED IDENTIFICATION

- VGC16 architecture



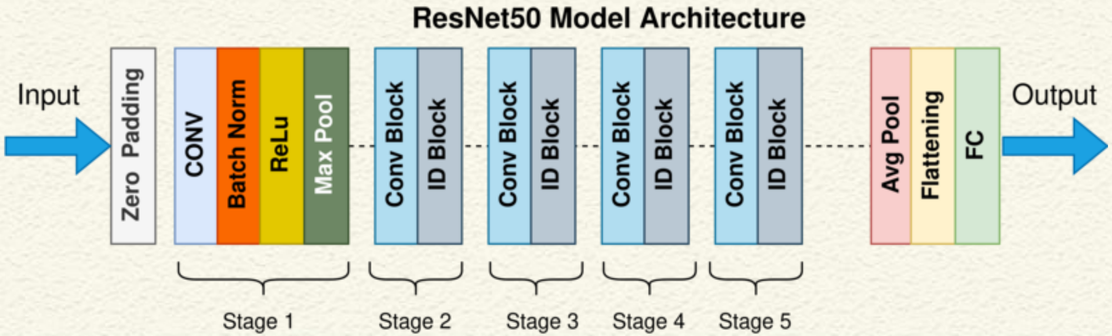
The first two layers have 64 channels of 3*3 filter size and same padding.

Output of 3rd fully connected layer is passed to softmax layer in order to normalize the classification vector.

Image source: Rohit Thakur, https://towardsdatascience.com/

CROP IMAGES + CNN FOR WEED IDENTIFICATION

- ResNet–50
  - Has a similar architecture as VGG16
  - Centered around a 3 × 3 convolutional layer with a ReLU activation function, but before and after each 3 × 3 convolutional layer 1 × 1 convolutional layers are established.
  - Furthermore, only one pooling layer is used, batch normalization is implemented, and the final total network structure comprised three times more layers than VGG16

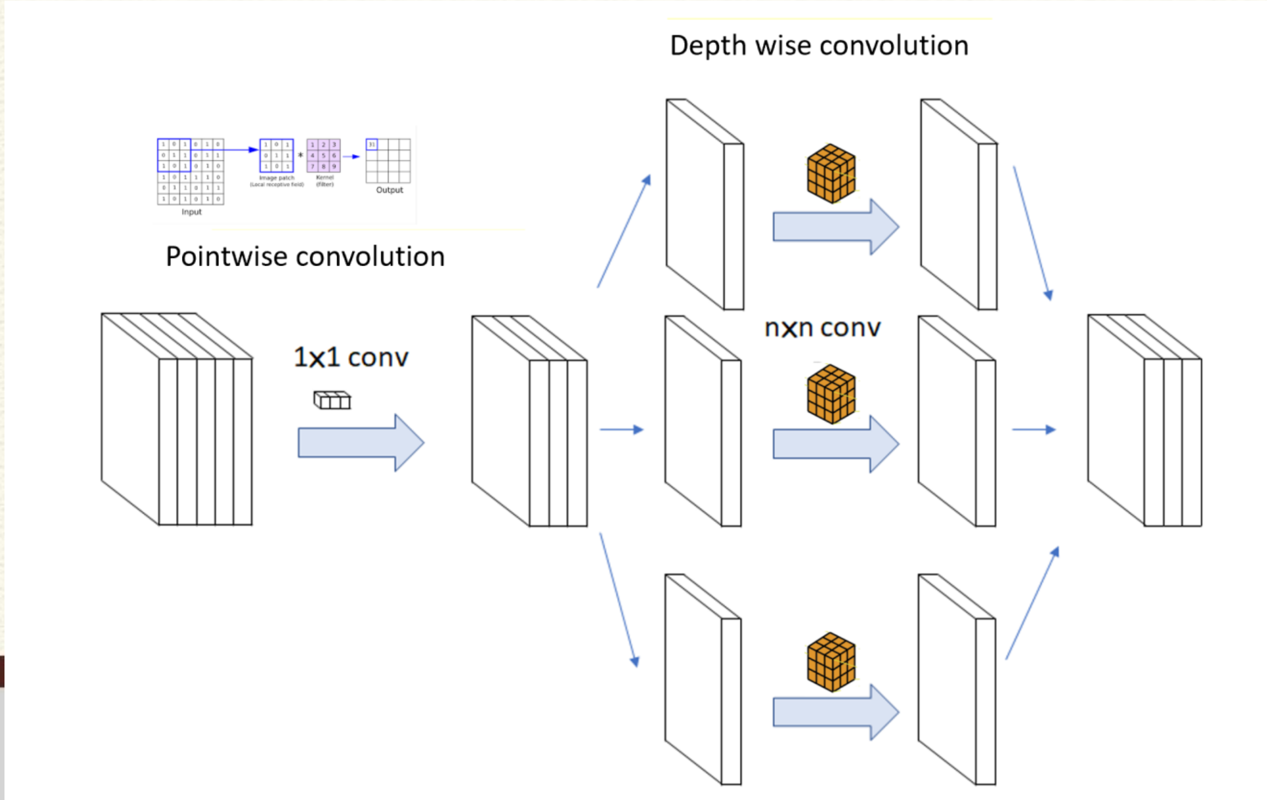


CROP IMAGES + CNN FOR WEED IDENTIFICATION

- Xception
  - Extreme version and an adaptation from Inception
  - While ResNet–50 tried to solve the image classification problem by increasing the depth of the network, the Inception architectures follow a different approach, by increasing the width of the network
  - A generic Inception module tries to calculate multiple different layers over the same input map in parallel, cleverly merging their results into the output
  - Three different convolutional layers and one max pool layer are activated in parallel, generating a wider CNN compared with the previous networks. Each output is then combined in a single concatenate layer
  - In Xception, the inception modules have been replaced with depth-wise separable convolutions

CROP IMAGES + CNN FOR WEED IDENTIFICATION

- **Xception**



CROP IMAGES + CNN FOR WEED IDENTIFICATION

- Data normalization and network training
    - Total 93130 images
    - 70% of the images were used for the training of the networks
    - 15% of the dataset was used for the validation performed in each training
    - The remaining 15% consisted of our testing subset, which was used for the final measurements and demonstration of the achieved results
    - The network experimentation was performed with Keras 2.4.3 in python 3.6.8 using the Tensorflow (2.3.0) backend
    - Transfer learning was used

CROP IMAGES + CNN FOR WEED IDENTIFICATION

- Transfer learning is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.

- For example, knowledge gained while learning to recognize cars could apply when trying to recognize truck

CROP IMAGES + CNN FOR WEED IDENTIFICATION

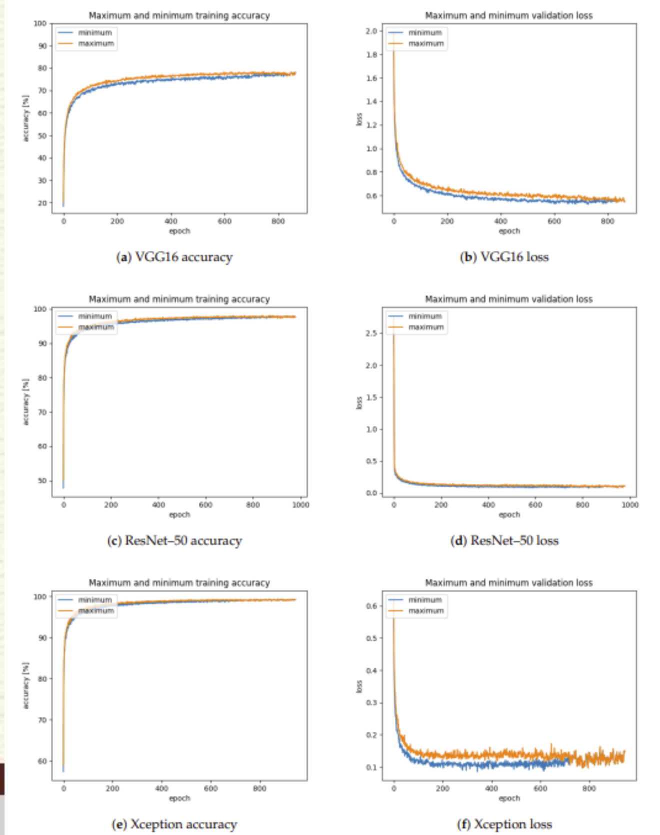
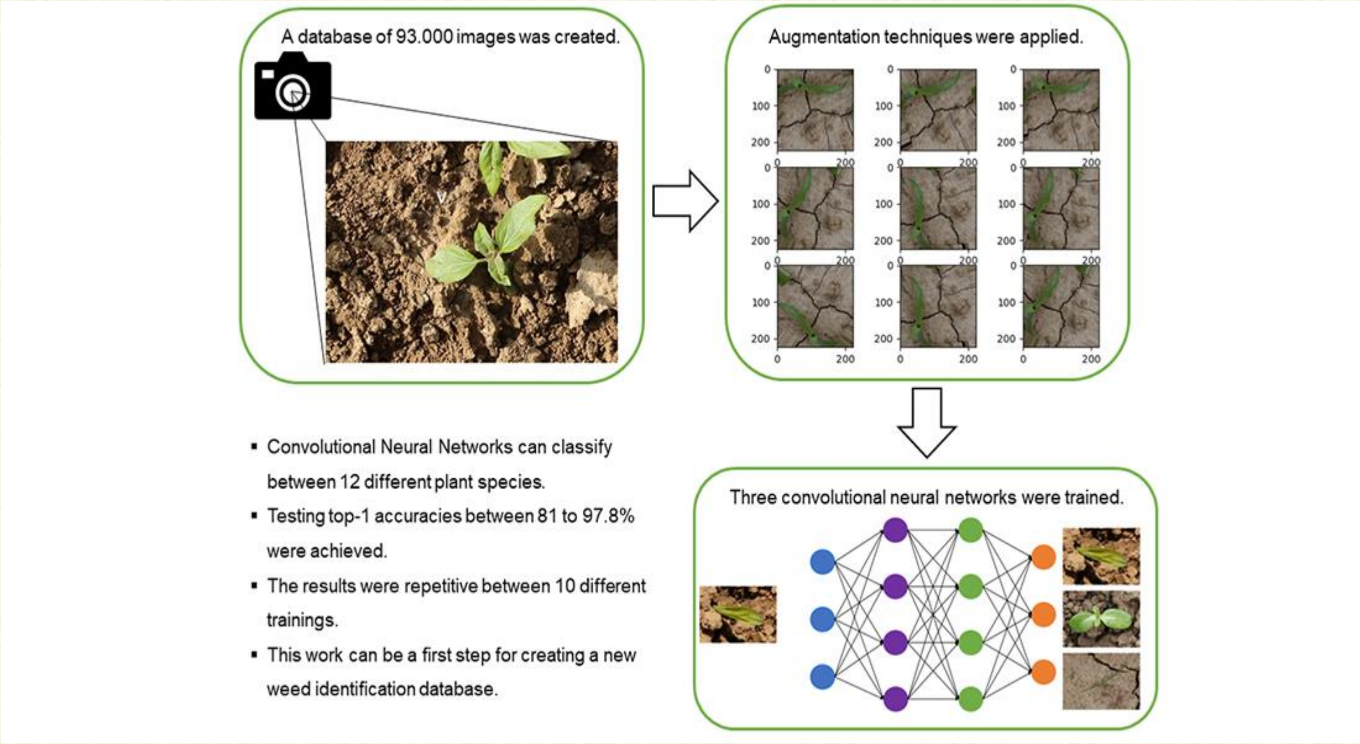


Figure 4. Minimum and maximum training and validation accuracy (**a**,**c**,**e**) along with the respective training and validation loss (**b**,**d**,**f**) over the ten repetitions performed for (**a**,**b**) VGG16, (**c**,**d**) ResNet–50, and the (**e**,**f**) Xception Convolutional Neural Networks

An epoch is a term used in machine learning and indicates the number of passes of the entire training dataset the machine learning algorithm has completed

CROP IMAGES + CNN FOR WEED IDENTIFICATION



REFERENCES

- Gómez-Robledo, L., López-Ruiz, N., Melgosa, M., Palma, A. J., Capitán-Vallvey, L. F., & Sánchez-Marañón, M. (2013). Using the mobile phone as Munsell soil-colour sensor: An experiment under controlled illumination conditions. Computers and electronics in agriculture, 99, 200-208.

- Peteinatos, G.G.; Reichel, P.; Karouta, J.; Andújar, D.; Gerhards, R. Weed Identification in Maize, Sunflower, and Potatoes with the Aid of Convolutional Neural Networks. Remote Sens. 2020, 12, 4185. https://doi.org/10.3390/rs12244185


Thank you

NPTEL ONLINE CERTIFICATION COURSES

Machine Learning for Soil and Crop Management

Prof. Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology Kharagpur

Week 8: UAV AND ML APPLICATIONS IN AGRICULTURE

LECTURE 37

CONCEPTS COVERED

- UAV
- UAV PARTS
- USE OF UAV FOR AGRICULTURAL OPERATIONS
- AGRICULTURAL BENEFITS OF DRONE

KEYWORDS

- UAV
- Payload
- Spraying
- Crop health
- Multispectral sensor

WHAT IS AN UAV?

- Unmanned aerial vehicle (UAV)
- Known as a drone
- An aircraft without any human pilot, crew, or passengers on board
- Unmanned aircraft system (UAS)= UAV + ground-based controller and a system of communications

APPLICATIONS OF UAV

- Defense operations
- Aerial photography
- Product deliveries
- Agriculture
- Policing and surveillance
- Infrastructure inspections

UAV IN AGRICULTURE

- Drones help farmers by precision agriculture
- Reduce the human error and increases efficiency in traditional farming methods
- Gathers data to regulate crop health, crop treatment, irrigation, crop damage assessment, and soil analysis
- Drone survey: minimal cost and time but boosts crop yield
- Digital sky platform= Drone + AIML + Remote Sensing

COMPONENTS OF DRONE

1. The aerial platform: the airframe, the navigation system, the power system, and the payload

2. The ground control station (GCS), which allows the human control from a remote emplacement

3. The communication system

PAYLOAD

Briefly, the payload is the weight a drone or unmanned aerial vehicle (UAV) can carry. It is usually counted outside of the weight of the drone itself and includes anything additional to the drone – such as extra cameras, sensors, or packages for delivery.

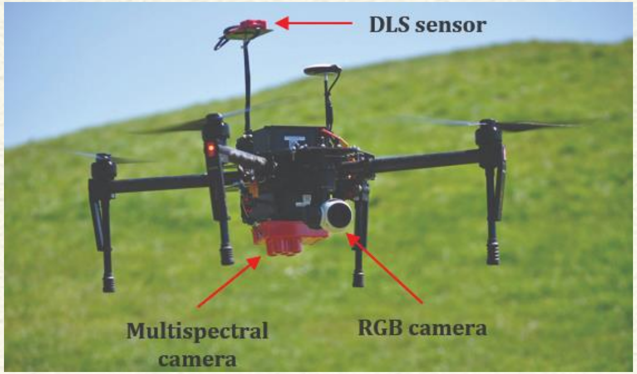
DRONE: DIFFERENT PARTS

1. Navigation system
2. GPS
3. Multiple sensors
4. Cameras
5. Programmable controllers
6. Tools for autonomous drones

Example: DJI

Phantom 4 RTK (Field Mapping)

DRONE SENSORS



Source: https://extensionpublications.unl.edu/assets/html/g2296/build/g2296.htm

DRONE TYPES

Fixed wing

Rotary wing

Rotary wing drones have multiple rotors with rotating blades. Drones with four rotors (quadcopters) and six rotors (hexcopters) are most common. Rotary wing drones allow for vertical takeoff, hovering, and closer crop inspection. They are easier to control manually than fixed wing drones. Generally, rotary wing drones are less expensive than fixed wing drones.

STEPS OF DATA COLLECTION BY A DRONE FOR AGRICULTURE

1. Analyzing the area
   i. Identify the area of operation
   ii. Establish a boundary
   iii. Analyze the area
   iv. Upload the GPS points into the Drone's navigation system
2. Using Autonomous drones: entering flight patterns into their system

STEPS OF DATA COLLECTION BY A DRONE FOR AGRICULTURE

3. Uploading the data

   i. Capturing data via sensors

   ii. Processing via softwares for analysis and interpretation

4. Output

   i. Formatting for easy understanding of the farmers

   ii. 3D mapping/ Photogrammetry

   iii. Vegetation index (Normalized Difference Vegetation Index (NDVI), Normalized Difference Red Edge (NDRE))

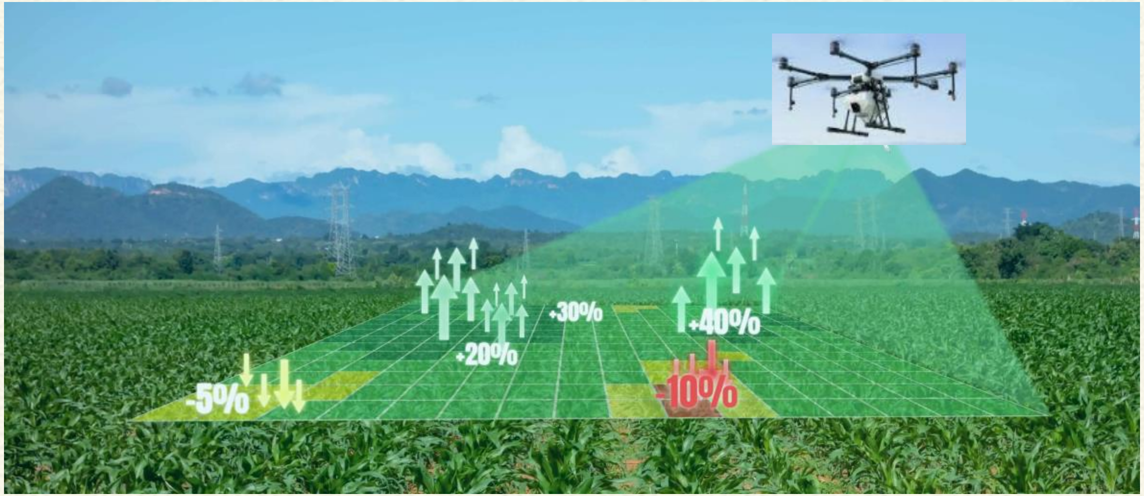
DRONE: IRRIGATION MONITORING

- Hyperspectral/ Multispectral/ Thermal
- Identify the areas that are too dry or need improvement by the farmer
- Drone survey: improve water use efficiency
- Identify the potential leaks/pooling in irrigation
- Calculate vegetation index: identify the healthy crops

DRONE: CROP HEALTH MONITORING

- Can identify the bacterial/fungal attack in crops
- Variable reflection of green light and NIRS: multispectral images to track crop health
- Quick monitoring the problem: save crop
- Crop failure: document the image for insurance claim

DRONE: CROP DAMAGE ASSESSMENT



- **Detect weeds by MS and RGB sensors**

- **Identify pests and infections**

- **Farmers can quantify the exact amount of chemicals needed**

DRONE: FIELD SOIL ANALYSIS

- Drone survey: obtain information about their land's soil conditions
- Multispectral sensors: allow seizing data useful for seed planting patterns, thorough field soil analysis, irrigation, and nitrogen-level management
- Precise Photogrammetry/ 3D mapping permits farmers to analyze their soil conditions thoroughly

DRONE: PLANTING

- Drone startups in India have invented drone-planting systems that allow drones to shoot pods, their seeds, and crucial nutrients into the soil
- Reduce costs by almost 85% but also increases consistency and efficiency

DRONE: SPRAYING AND LIVESTOCK TRACKING

- Spraying drone reduces human contact with harmful chemicals
- Agri-drone: quick spraying
- Monitor the movements of cattle
- Thermal sensor technology helps find lost animals and detect an injury or sickness

DRONE: SPRAYING AND LIVESTOCK TRACKING

BENEFITS OF DRONE USE IN AGRICULTURE

- Enhanced production: comprehensive irrigation planning and monitoring crop health and soil health
- Effective and adaptive techniques based on weather condition and allocate resource without wastage
- Greater safety of the farmers: spraying pesticides in challenging terrain, infected areas, taller crops and power lines

BENEFITS OF DRONE USE IN AGRICULTURE

- Quick operation: quick and mindful decisions without second-guessing. Various sensors of the drone enable capturing and analyzing data from the entire field.
- Less wastage of resources - Agri-drones enables optimum usage of all resources such as fertilizer, water, seeds, and pesticides
- 99% Accuracy rate - The drone survey helps farmers calculate the precise land size, segment the various crops, and indulge in soil mapping

BENEFITS OF DRONE USE IN AGRICULTURE

- Useful for Insurance claims - Farmers use the data or images captured through drones to claim crop insurance in case of any crop damages. They even calculate risks/losses associated with the land while being insured.
- Evidence for insurance companies - agricultural insurance sectors use Agri-drones for efficient and trustworthy data. They capture the photos of crop damages that have occurred for the right estimation of monetary payback to the farmers
- Fight locust swarm

REFERENCES

- https://www.equinoxsdrones.com/blog/importance-of-drone-technology-in-indian-agriculture-farming
- https://extensionpublications.unl.edu/assets/html/g2296/build/g2296.htm


Thank you

NPTEL ONLINE CERTIFICATION COURSES

Machine Learning for Soil and Crop Management

Prof. Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology Kharagpur

Week 8: UAV AND ML APPLICATIONS IN AGRICULTURE

LECTURE 38

CONCEPTS COVERED

- UAV FOR DIFFERENT USES
- RECURRENT NEURAL NETWORKS

KEYWORDS

- UAV
- Payload
- RNN
- BPTT
- Feedforward networks

COMPARISON BETWEEN MULTIROTOR AND FIXED-WING UAV

Type	Pay-load	Flight time	Benefits	Limitation	Examples
Multirotor UAV	0.8–8.0 kg	8 – 120 min	• Applicable with waypoint navigation • Hovering capabilities • Can hold range of sensors from thermal, multispectral to hyperspectral cameras	• Payload may limit battery usage and flight time	DJI Inspire, Mikrocopter ARK OktoXL 6S12, Yamaha RMAX
Fixed wing UAV	1.0–10 lg	30 – 240 min	• Better flight time • Multiple sensors can be mounted • Limited hovering capacity	• Lower speed is required for image stitching	Landcaster Precision Hawk, senseFly eBee

Norasma et al. (2019), Unmanned Aerial Vehicle Applications In AgricultureIOP Conf. Series: Materials Science and Engineering 506: 012063
doi:10.1088/1757-899X/506/1/012063

PLANT PHENOTYPING

UAV and Sensor type	Applications	Photo
3DR Iris+ (a) and DJI Phantom 2 (b), camera gimbal and GoPro camera	Agriculture field monitoring, autonomous navigation	
Turnigy 9XR Octocopter UAV, Digital camera (RGB)	Basal Stem Rot (BSR) disease in oil palm	
Sensefly eBee UAV, 16-megapixel digital camera	Mapping changes to land cover, transmission of infectious diseases	
(HiSystems GmbH Mikrokopter, Germany, RGB camera	Develop a new estimation technique for disease severity	
Microdrones MD4-200, A Tetracam ADC Lite digital camera	NDVI and grain yield, aerial biomass and nitrogen content	

Norasma et al. (2019), Unmanned Aerial Vehicle Applications In AgricultureIOP Conf. Series: Materials Science and Engineering 506: 012063 doi:10.1088/1757-899X/506/1/012063

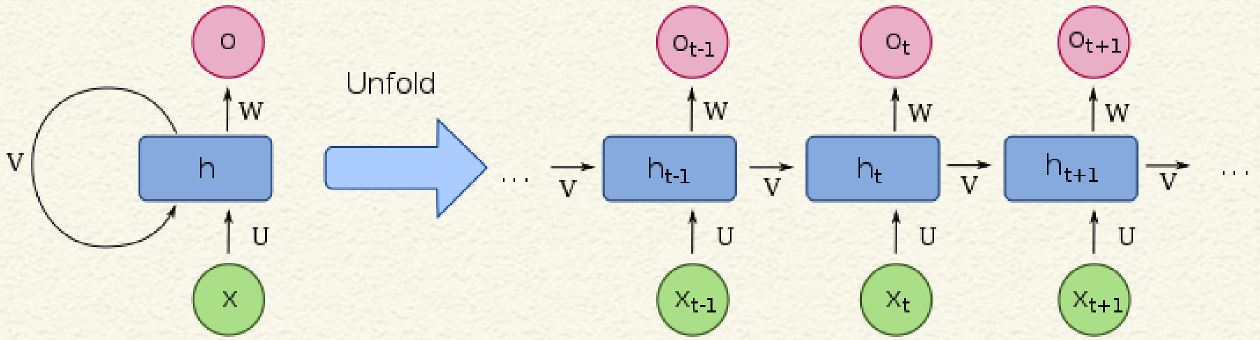
RECURRENT NEURAL NETWORK (RNN)

- A type of ANN
- Uses sequential data or time series data
- Commonly used for ordinal or temporal problems, such as
  - Language translation
  - Natural language processing (nlp)
  - Speech recognition
  - Image captioning

RNN APPLICATIONS

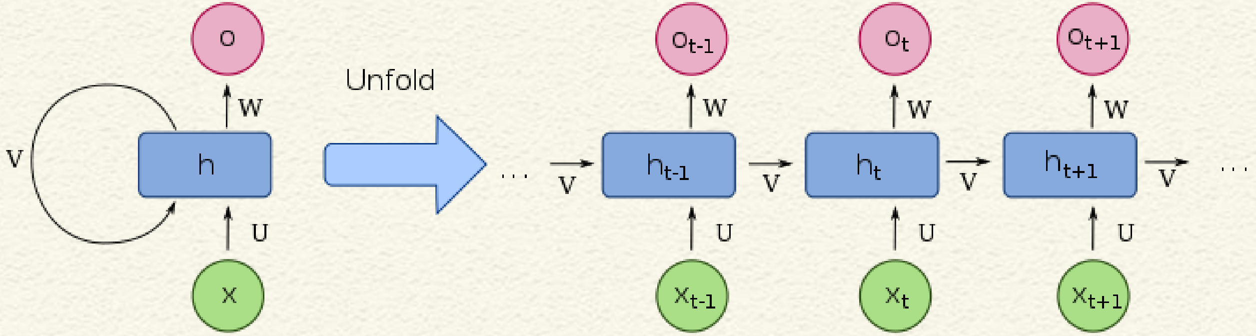
RECURRENT NEURAL NETWORK (RNN)

- Traditional ANN: assumes independence of inputs and outputs
- RNN output: depend on the prior elements within the sequence
- Utilize memory to control the current input or output

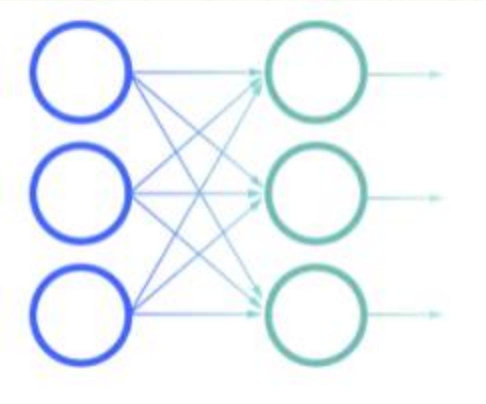


RECURRENT NEURAL NETWORK (RNN)

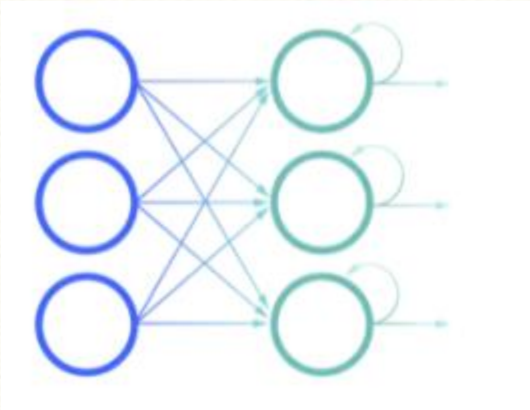
- Output from previous step are fed as input to the current step
- RNN: when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words



RECURRENT NEURAL NETWORK (RNN)



Feedforward Neural Networks



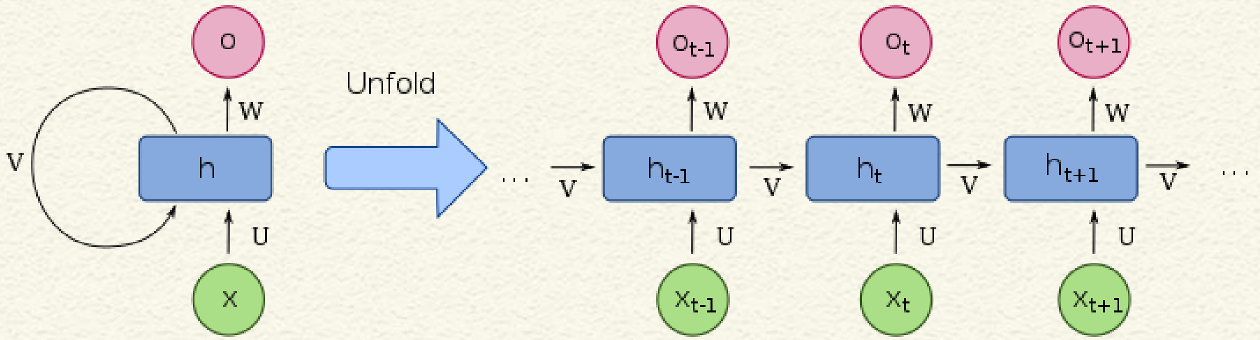
RNN

RECURRENT NEURAL NETWORK (RNN): HOW IT WORKS?

- Idiom example: sit on the fence (to remain neutral)
- It needs to be expressed in that specific order to make sense
- RNN: consider the position of each word and use this info for predict the next word in the sequence

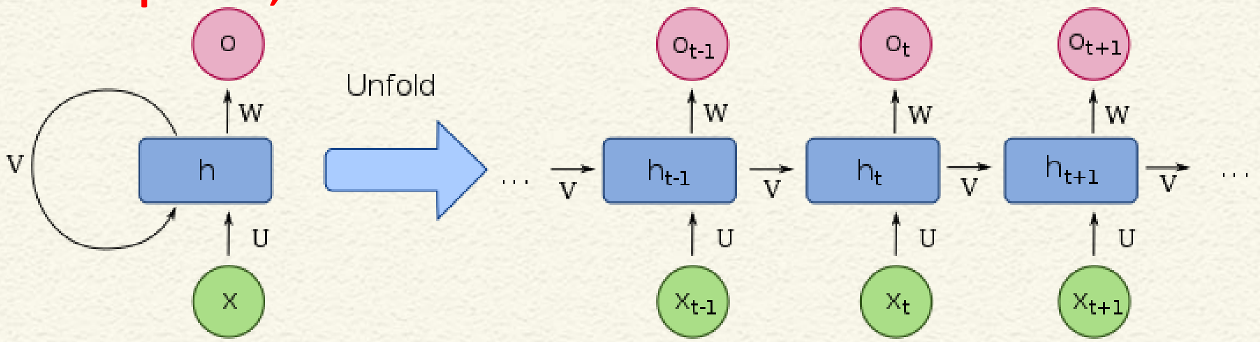
RECURRENT NEURAL NETWORK (RNN)

- Compressed visual: RNN represents the whole neural network and predict the entire phrase: sit on the fence
- Unfold visual: represents individual layers/time steps of NN



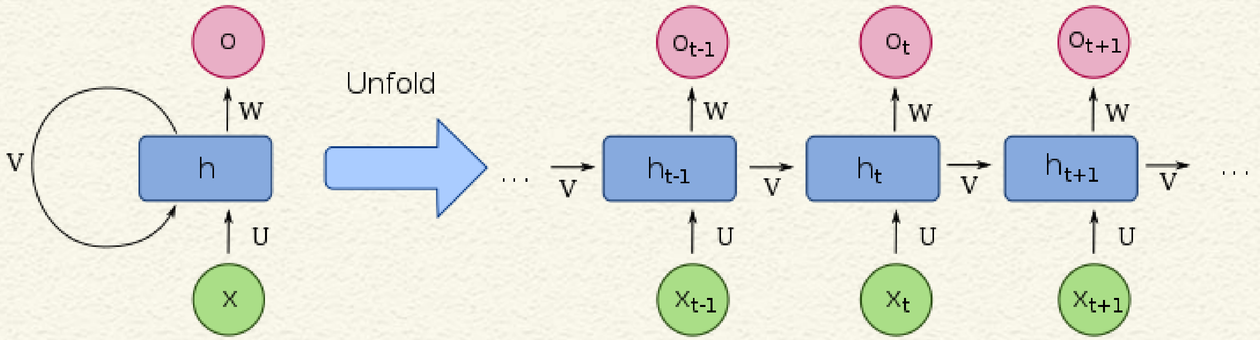
RECURRENT NEURAL NETWORK (RNN)

- Each layer maps to a single word in that phrase
- Prior inputs, such as “sit” and “on”, would be represented as a hidden state in the third timestep to predict the output in the sequence, “the”.



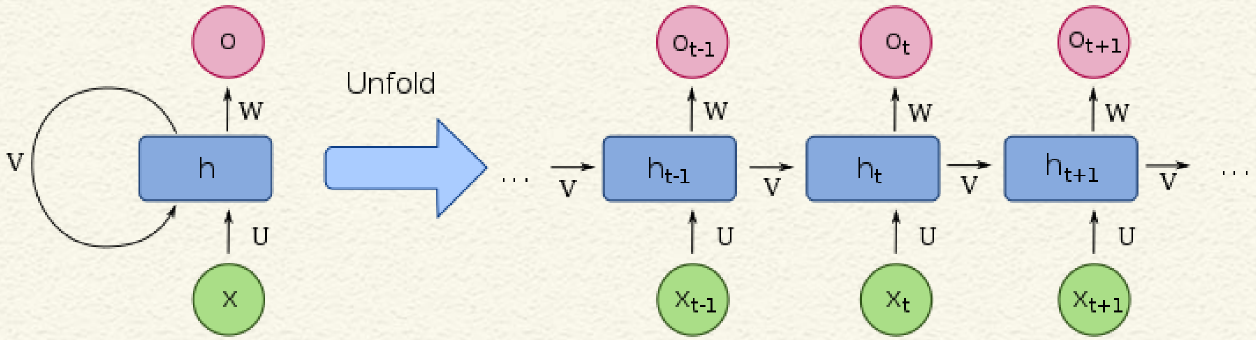
RECURRENT NEURAL NETWORK (RNN)

- Feedforward networks have different weights across each node
- **RNN:** share the same weight parameter within each layer of the network (adjusted by backpropagation)



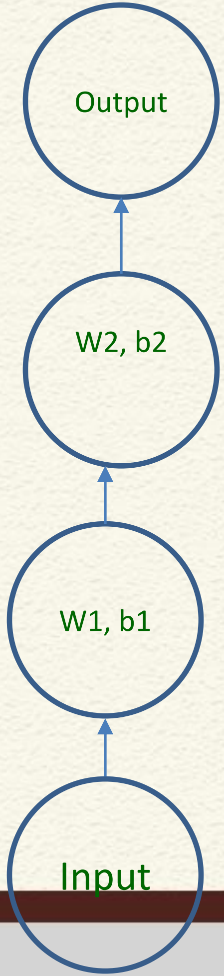
RECURRENT NEURAL NETWORK (RNN)

- **RNN: share the same weight parameter within each layer of the network (adjusted by backpropagation).**
- **WHY?????? To reduces the complexity of parameters**



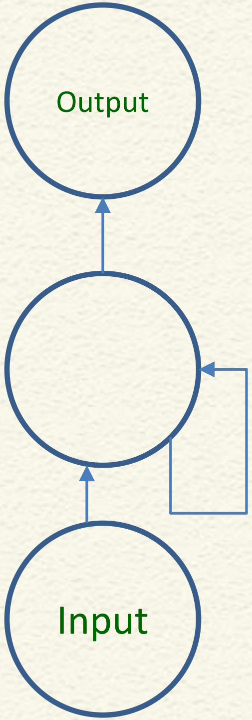
RECURRENT NEURAL NETWORK (RNN)

- A deeper network with one input layer, two hidden layers and one output layer.
- Each hidden layer : own set of weights and biases
- Each of these layers are independent of each other, i.e. they do not memorize the previous outputs.



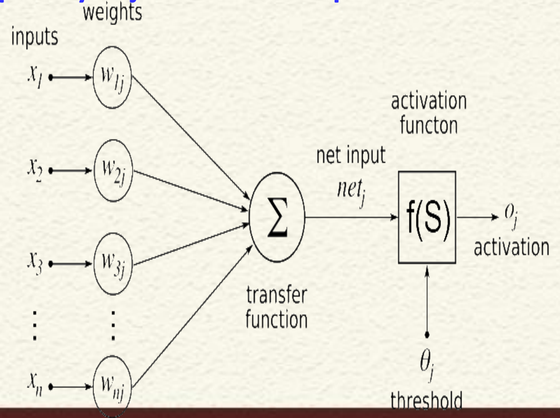
RECURRENT NEURAL NETWORK (RNN)

- RNN: converts the independent activations into dependent activations by providing the same weights and biases to all the layers, thus reducing the complexity of increasing parameters and memorizing each previous outputs by giving each output as input to the next hidden layer
- Hence these two layers can be joined together such that the weights and bias of all the hidden layers is the same, into a single recurrent layer



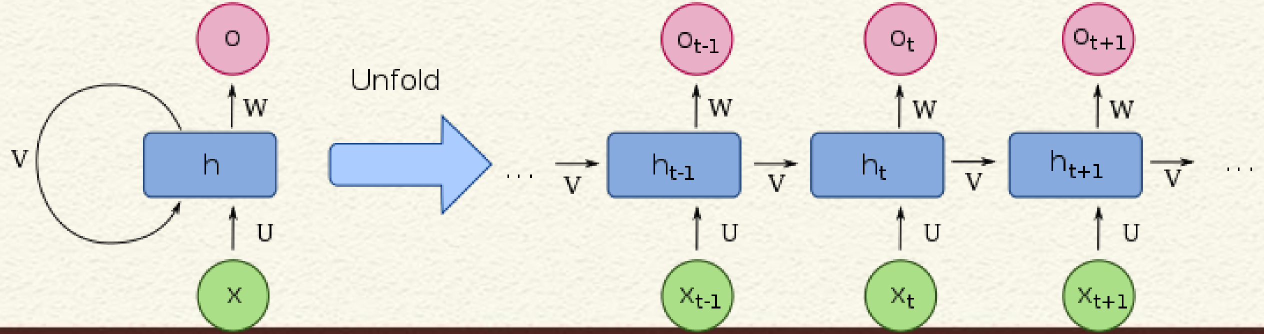
RECALL: ANN

- Most deep neural networks are feedforward, meaning they flow in one direction only, from input to output.
- Backpropagation:
  - Move in the opposite direction from output to input
  - Calculate and attribute the error associated with each neuron, subsequently adjust and fit the parameters of the model(s)



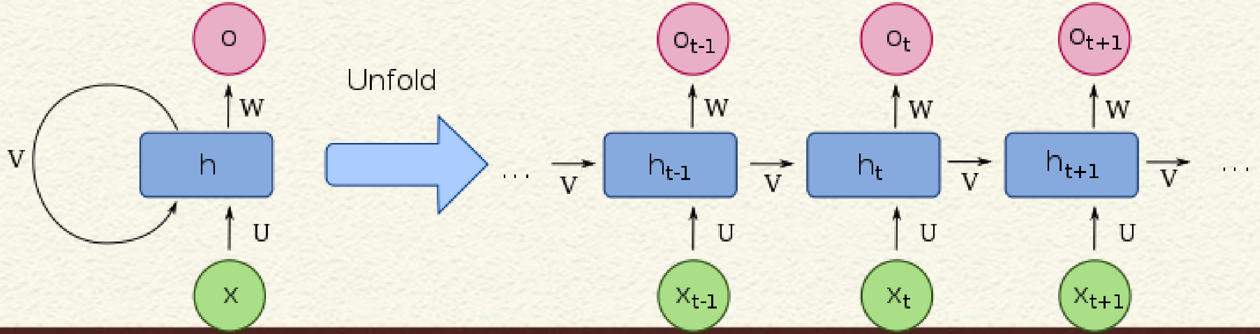
RECURRENT NEURAL NETWORK (RNN)

- RNN leverage backpropagation through time (BPTT) algorithm to determine the gradients
- Slightly different than traditional ANN
- Specific to sequence data



RECURRENT NEURAL NETWORK (RNN)

- **BPTT differs from the traditional approach in that BPTT sums errors at each time step whereas feedforward networks do not need to sum errors as <u>they do not share parameters across each layer</u>**



REFERENCES

https://www.ibm.com/cloud/learn/recurrent-neural-networks

Norasma et al. (2019), Unmanned Aerial Vehicle Applications In AgricultureIOP Conf. Series: Materials Science and Engineering 506: 012063 doi:10.1088/1757-899X/506/1/012063


Thank you

NPTEL ONLINE CERTIFICATION COURSES

Machine Learning for Soil and Crop Management

Prof. Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology Kharagpur

Week 8: UAV AND ML APPLICATIONS IN AGRICULTURE

LECTURE 39

CONCEPTS COVERED

- TYPES OF RNN
- TRAINING RNN
- ADVANTAGES AND DISADVANTAGES OF RNN
- VSION TRANSFORMER

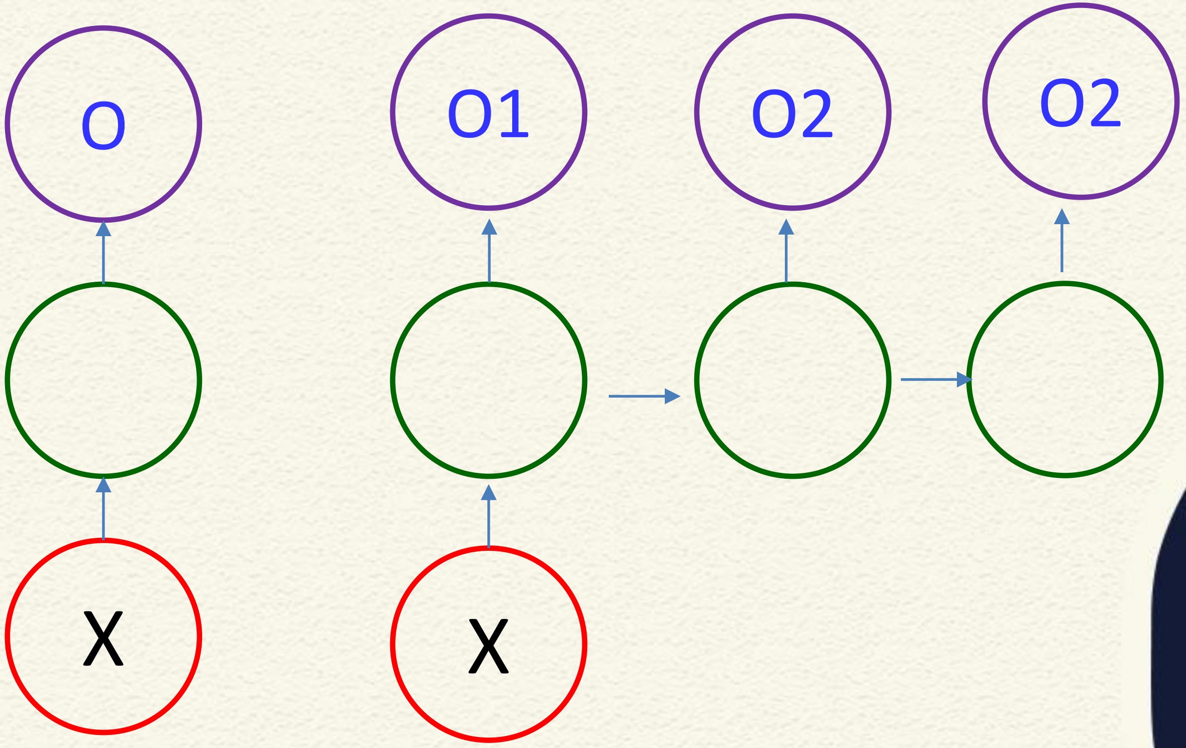
KEYWORDS

- Bidirectional recurrent neural networks
- Long Short Term Memory
- Vision Transformer
- Attention
- Self attention

TYPES OF RNN

- Feedforward networks map one input to one output
- RNN do not actually have this constraint
- RNN inputs and outputs can vary in length, and different types of RNNs are used for different use cases, such as music generation and machine translation

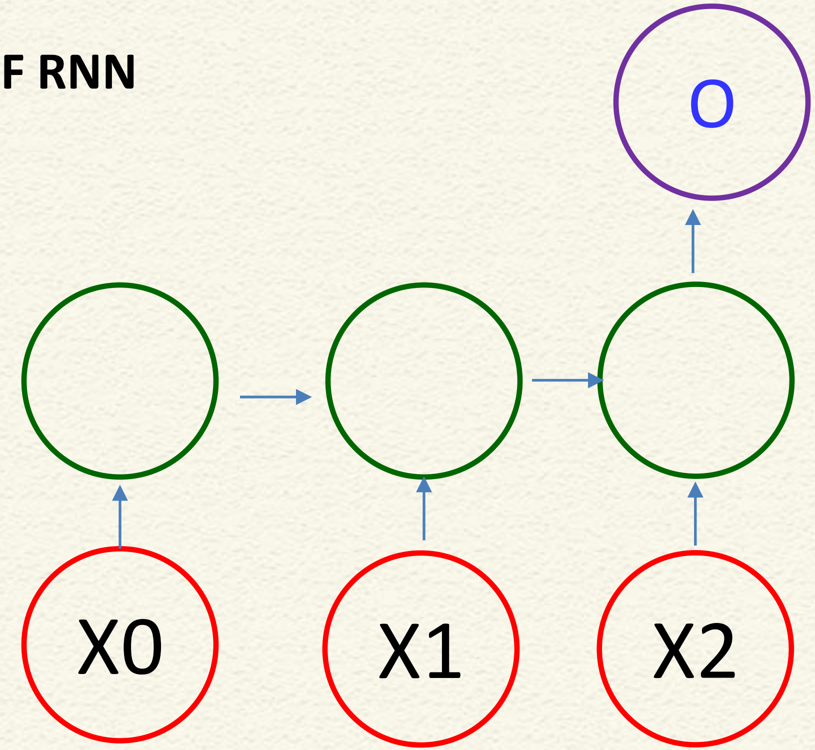
TYPES OF RNN



One-to-one

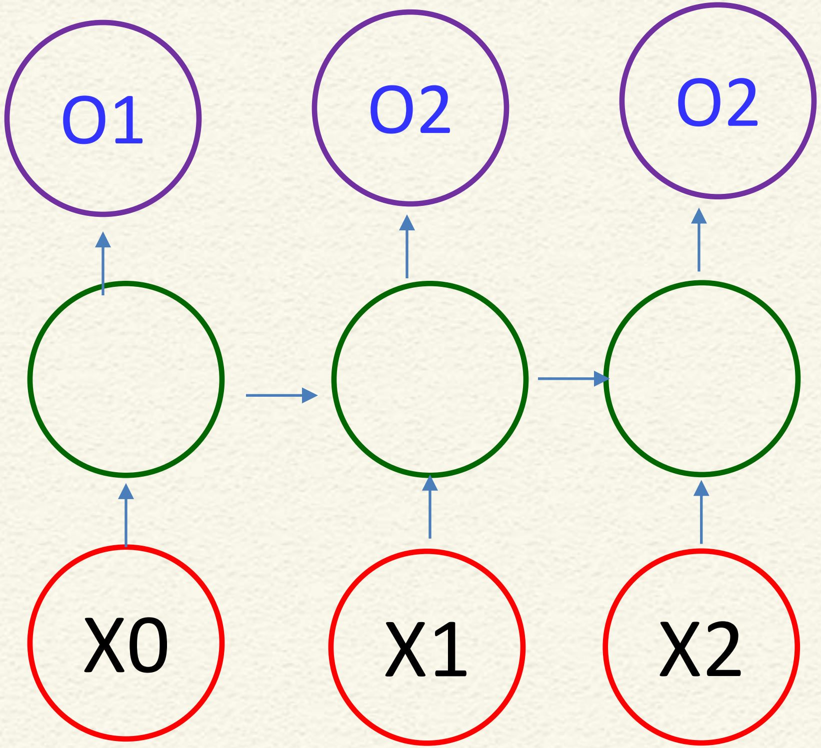
One-to-many

TYPES OF RNN



Many to one

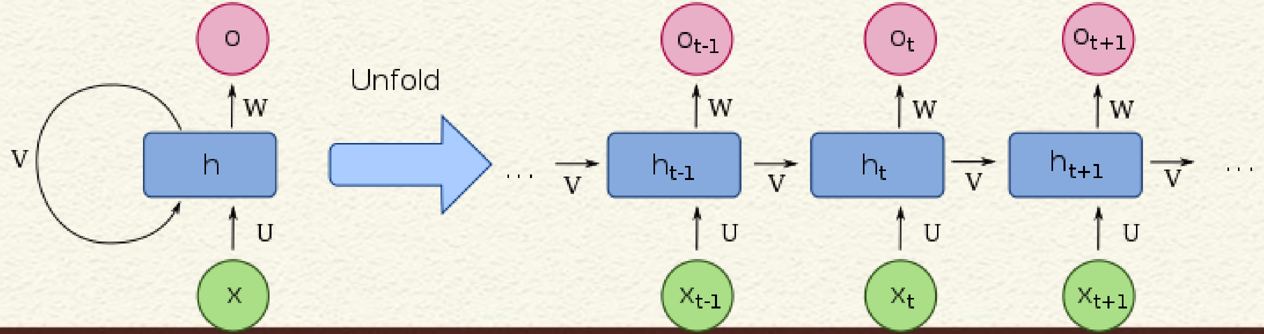
TYPES OF RNN



Many to many

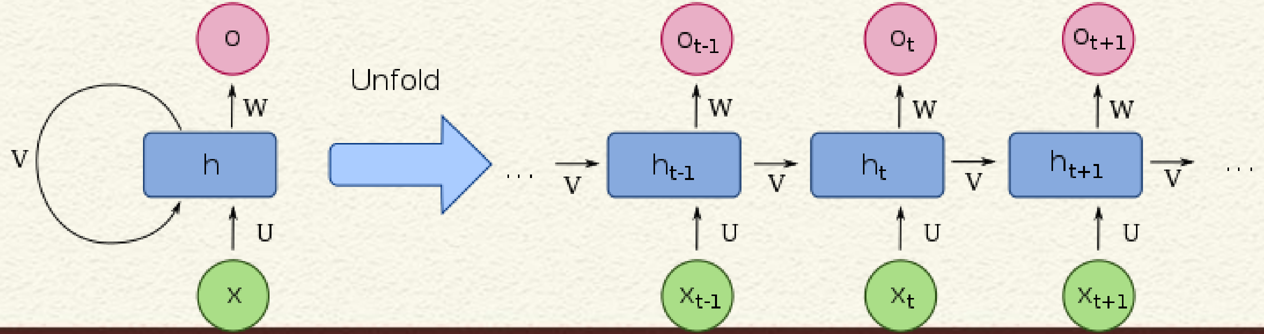
TRAINING A RECURRENT NEURAL NETWORK (RNN)

- Step 1: A single time step of the input is provided to the network
- Step 2: Then calculate its current state using set of current input and the previous state
- Step 3: The current ht becomes ht-1 for the next time step



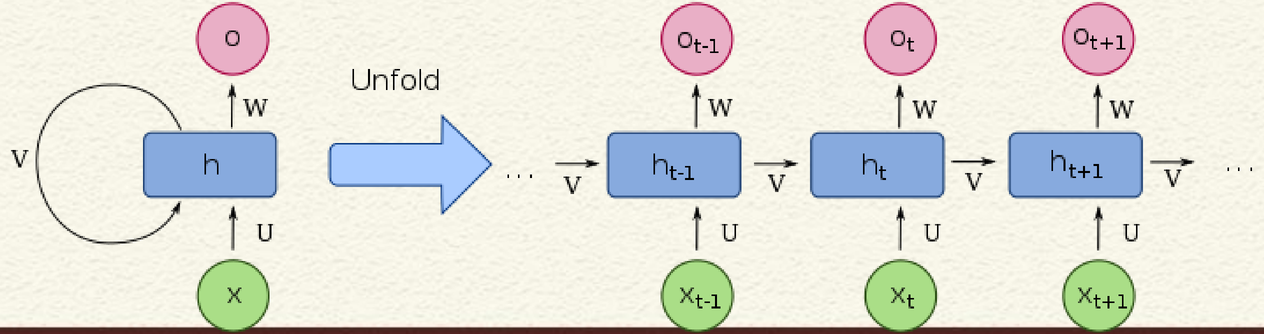
TRAINING A RECURRENT NEURAL NETWORK (RNN)

- **Step 4: One can go as many time steps according to the problem and join the information from all the previous states**
- **Step 5: Once all the time steps are completed the final current state is used to calculate the output**



TRAINING A RECURRENT NEURAL NETWORK (RNN)

- Step 6: The output is then compared to the actual output i.e the target output and the error is generated
- Step 7: The error is then back-propagated to the network to update the weights and hence the network (RNN) is trained



VARIANT OF RNN

Bidirectional recurrent neural networks (BRNN):

- A variant network architecture of RNNs
- While unidirectional RNNs can only drawn from previous inputs to make predictions about the current state
- BRNNs pull in future data to improve the accuracy of it
- If we return to the example of "sit on the fence" earlier, the model can better predict that the second word in that phrase is "on" if it knew that the last word in the sequence is "fence."

ADVANTAGES AND DISADVANTAGES OF RNN

ADVANTAGES

- An RNN remembers each and every information through time. It is useful in time series prediction only because of the feature to remember previous inputs as well. This is called Long Short Term Memory
- RNN are even used with convolutional layers to extend the effective pixel neighborhood

ADVANTAGES AND DISADVANTAGES OF RNN

DISADVANTAGES

- Training an RNN is a very difficult task
- It cannot process very long sequences if using ReLu as an activation function

UAV IMAGES + DL FOR WEED AND CROP CLASSIFICATION

- Monitoring crops and weeds is a major challenge in agriculture and food production
- Weeds compete directly with crops for moisture, nutrients, and sunlight
- Weed detection and mapping: an essential step in weed control
- Deep learning approaches have shown good performance in many agriculture-related remote sensing tasks, such as plant classification, disease detection, etc.
- Challenges : high computation cost, the need of large labelled datasets, intra-class discrimination (in growing phase weeds and crops share many attributes similarity as color, texture, and shape).

UAV IMAGES + DL FOR WEED AND CROP CLASSIFICATION

- Automatically recognize weeds and crops in drone images using the vision transformer approach
- The main objective was to study the paradigm of transformers architectures for specific tasks such as plant recognition in UAV images, where labeled data are not available in large quantities.
- Data augmentation and transfer learning were used as a strategy to fill the gap of labeled data

Reedha et al. (2022)

CONTRIBUTIONS

1. Low-altitude aerial imagery based on UAVs and self-attention algorithms for crop management.
2. First study to explore the potential of transformers for classification of weed and crop images.
3. Evaluation of the generalization capabilities of deep learning algorithms with regard to train set reduction, in crop plants classification task.

VISION TRANSFORMER (ViT)

The Vision Transformer, or ViT, is a model for image classification that employs a Transformer-like architecture over patches of the image

An image is split into fixed-size patches, each of them are then linearly embedded, position embeddings are added, and the resulting sequence of vectors is fed to a standard Transformer encoder

In order to perform classification, the standard approach of adding an extra learnable "classification token" to the sequence is used

Credit: David Coccomini (CC BY-SA 4.0)

TRANSFORMER

- A deep learning model that adopts the mechanism of self-attention, differentially weighting the significance of each part of the input data
- Like RNN, transformers are designed to handle sequential input data, such as natural language, for tasks such as translation and text summarization
- However, unlike RNNs, transformers do not necessarily process the data in order. Rather, the attention mechanism provides context for any position in the input sequence

TRANSFORMER

- **For example, if the input data is a natural language sentence, the transformer does not need to process the beginning of the sentence before the end. Rather, it identifies the context that confers meaning to each word in the sentence. This feature allows for more parallelization than RNNs and therefore reduces training times**

Vision Transformers

Transformers | Davide Coccomini | 2021

Credit: David Coccomini (CC BY-SA 4.0)

SELF ATTENTION

- In neural networks, attention is a technique that mimics cognitive attention.
- The effect enhances some parts of the input data while diminishing other parts — the thought being that the network should devote more focus to that small but important part of the data

Credit: David Coccomini (CC BY-SA 4.0)

ViT ARCHITECTURE

1. Split an image into patches (fixed sizes)
2. Flatten the image patches
3. Create lower-dimensional linear embeddings from these flattened image patches
4. Include positional embeddings
5. Feed the sequence as an input to a state-of-the-art transformer encoder
6. Pre-train the ViT model with image labels, which is then fully supervised on a big dataset
7. Fine-tune on the downstream dataset for image classification

Vision Transformers

Transformers | Davide Coccomini | 2021

Credit: David Coccomini (CC BY-SA 4.0)

REFERENCES

- Reedha, R.; Dericquebourg, E.; Canals, R.; Hafiane, A.Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images. Remote Sens. 2022, 14, 592. https://doi.org/10.3390/rs14030592

- https://viso.ai/deep-learning/vision-transformer-vit/

Thank you

NPTEL ONLINE CERTIFICATION COURSES

Machine Learning for Soil and Crop Management

Prof. Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology Kharagpur

Week 8: UAV AND ML APPLICATIONS IN AGRICULTURE

LECTURE 40

CONCEPTS COVERED

- UAV BASED WEED AND CROP CLASSIFICATION

KEYWORDS

- Drone image
- Weed classification
- Self-attention
- F1-score
- Vision Transformer

CROP FIELDS, UAV, AND IMAGE SENSOR

1. Beet
2. Parsley
3. Spinach

(a) (b)

Figure 1. Apparatus used for data acquisition. (**a**) Starfury Drone; (**b**) Sony ILCE-7R Camera.

IMAGE ACQUISITION

1. 30 m = beet field
2. 20 m = parsley and spinach field
3. These altitudes where selected to minimize drone flight times while maintaining sufficient image quality

Figure 1. Apparatus used for data acquisition. (a) Starfury Drone; (b) Sony ILCE-7R Camera.

IMAGE ACQUISITION

- The drone followed a specific flight plan and the camera captured RGB images at regular intervals.
- The images captured have respectively a minimum longitudinal and lateral overlapping of 70% and 50–60% depending on the fields vegetation coverage and homogeneity, assuring a better and complete coverage of the whole field of 4 ha and improving the accuracy of the orthorectified image of the field.

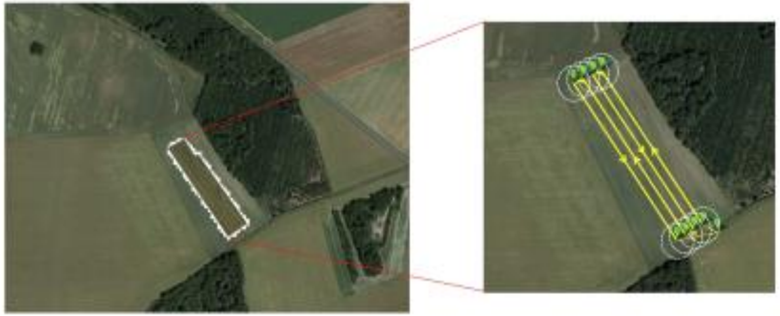
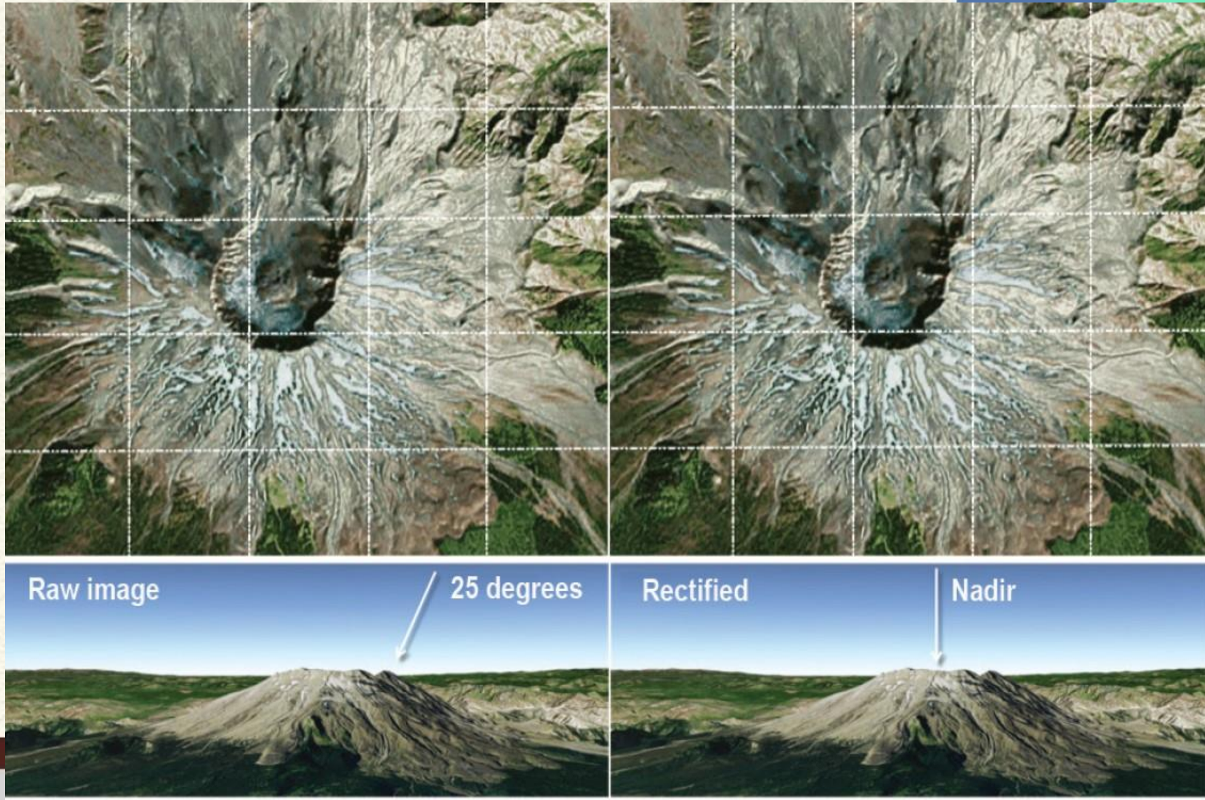


Figure 2. Overlay of the orthophoto on google earth of the spinach plot (**left**) and the flight plan (**right**) across a spinach field (the images are taken along the yellow lines at regular intervals to ensure sufficient overlapping).

Figure 3. Example of image captured from a spinach study site.

ORTHORECTIFIED IMAGE

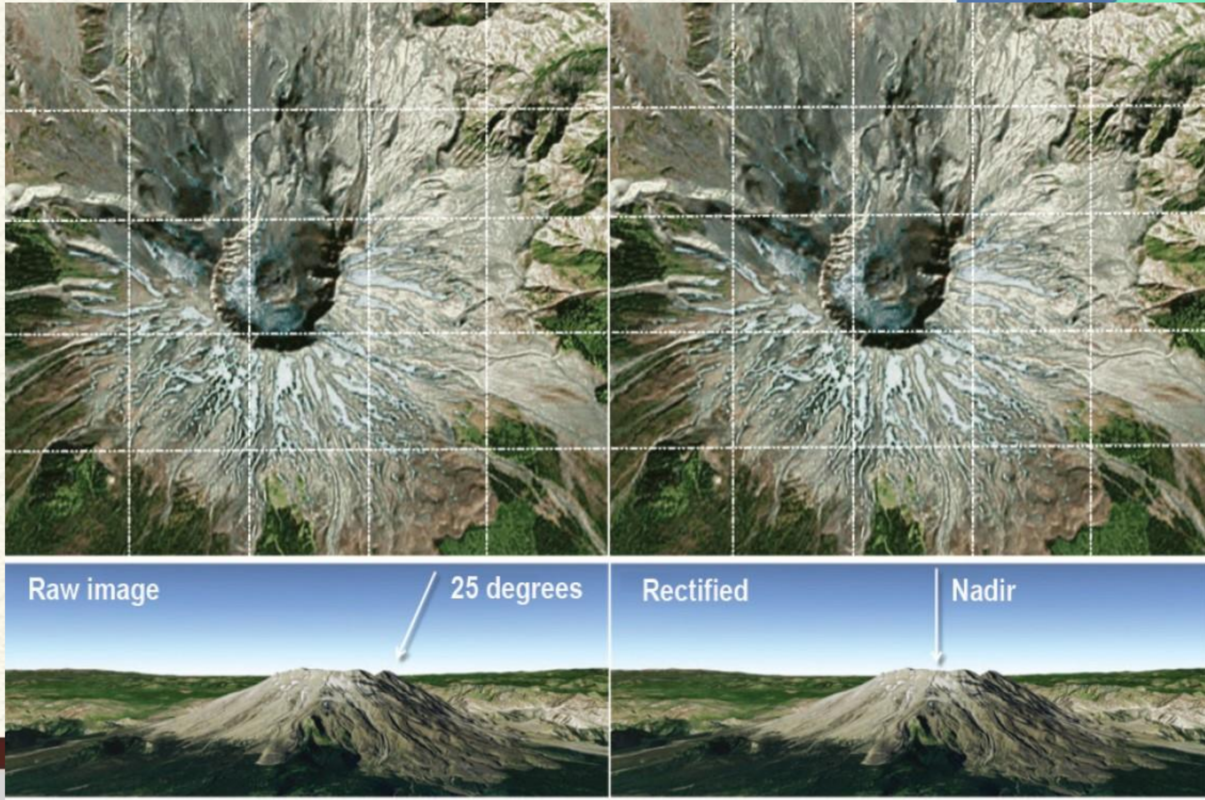
- Raw aerial or satellite imagery cannot be used in a GIS until it has been processed such that all pixels are in an accurate (x,y) position on the ground.
- Photogrammetry is a discipline, for processing imagery to generate accurately georeferenced images, referred to as orthorectified images (or sometimes simply orthoimages).
- Orthorectified images have been processed to apply corrections for optical distortions from the sensor system, and apparent changes in the position of ground objects caused by the perspective of the sensor view angle and ground terrain.



https://www.esri.com/about/newsroom/insider/what-is-orthorectified-imagery/

ORTHORECTIFIED IMAGE

- The orthorectification process requires: An accurate description of the sensor, typically called the sensor model; detailed information about the sensor location and orientation for every image; and an accurate terrain model



https://www.esri.com/about/newsroom/insider/what-is-orthorectified-imagery/

LABELLING THE IMAGES

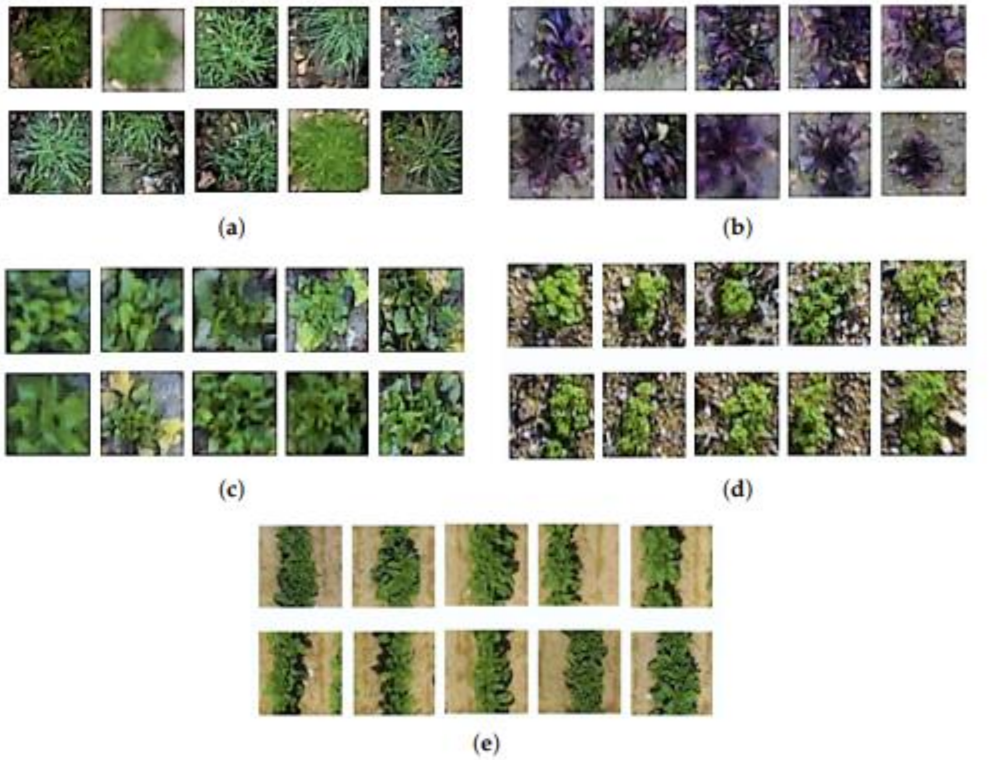


Figure 4. This overview shows sample images patches of all 5 classes of our custom dataset. The images measure 64 × 64 pixels. Each class contains 3200 to 4000 images. (**a**) Weeds; (**b**) Beet; (**c**) Off-type green leaves beet; (**d**) Parsley; (**e**) Spinach.

IMAGE PREPROCESSING

- Images have been rescaled to 0–1 range and then normalized by scaling the pixels values to have a zero mean and unit variance before being divided into training, validation and testing sets
- Data augmentation to improve the model robustness

Table 1. Class Distribution.

Class	Number
Weed	4000
Beet	4000
Off-type Beet	3265
Parsley	4000
Spinach	4000

REACLL: VISION TRANSFORMER (ViT)

- The Vision Transformer, or ViT, is a model for image classification that employs a Transformer-like architecture over patches of the image
- An image is split into fixed-size patches, each of them are then linearly embedded, position embeddings are added, and the resulting sequence of vectors is fed to a standard Transformer encoder
- In order to perform classification, the standard approach of adding an extra learnable "classification token" to the sequence is used
- The self-attention enhances some parts of the input data while diminishing other parts — the thought being that the network should devote more focus to that small but important part of the data

Vision Transformers

Transformers | Davide Coccomini | 2021

SELF-ATTENTION

- If each pixel in a feature map is regarded as a random variable and the covariances are calculated, the value of each predicted pixel can be enhanced or weakened based on its similarity to other pixels in the image
- The mechanism of employing similar pixels in training and prediction and ignoring dissimilar pixels is called the self-attention mechanism
- It helps to relate different positions of a single sequence of image patches in order to gain a more vivid representation of the whole image

Credit: David Coccomini (CC BY-SA 4.0)

SELF-ATTENTION

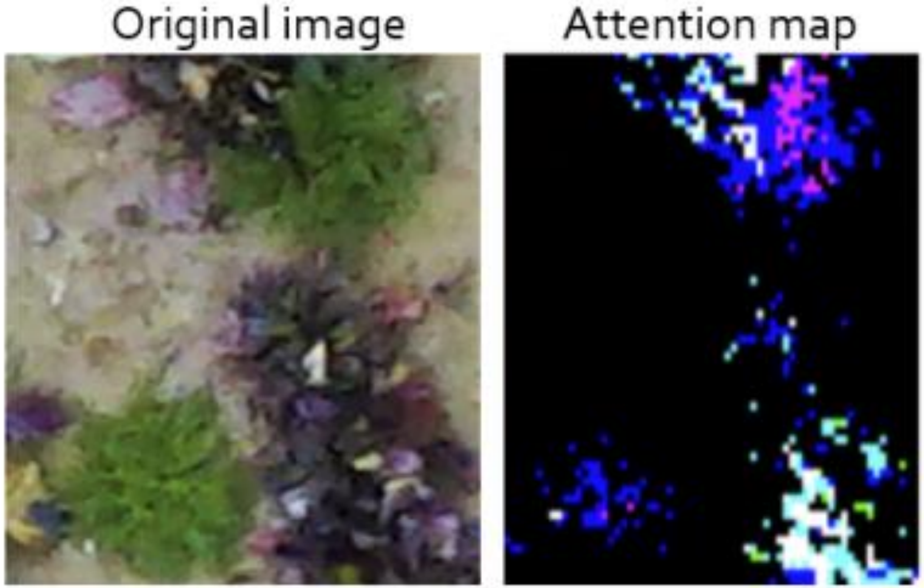


Figure 5. Attention mechanism on an image patch (**left**) containing weeds (in green) and beet plant (in red). With the original image on the left and the attention map (**right**) obtained with ViT-B16 model. The attention map shows the model's attention on the different plants: with a dark blue and purple colour pixel representing the attention on the weeds and a light blue colour pixel representing the beet plant.

SELF-ATTENTION

When patch 1 is passed through the transformer, self-attention will calculate how much attention should pay to others (patch 2, patch 3, . . .).

In addition, every head will have one attention pattern and finally, they will sum up all attention patterns (all heads).

The model tries to identify the object (weed) on the image and tries to focus its attention on it (as it stands out from the background)

Figure 6. Attention map generated from layers 7 to 12 of the ViT-B16 model on an image of a weed.

ViT

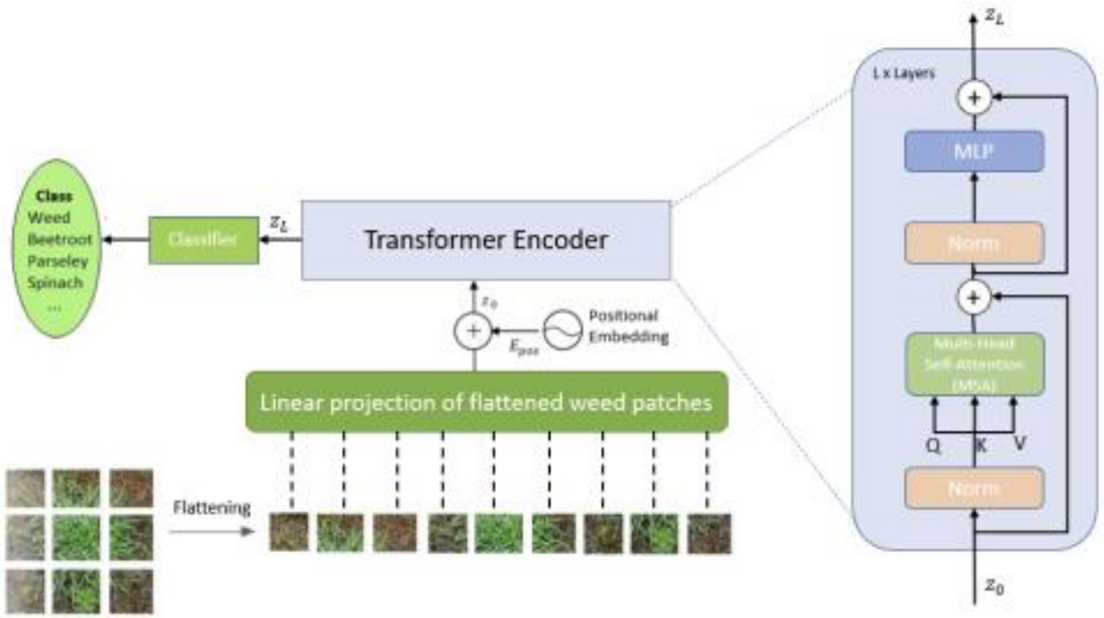


Figure 7. ViT model architecture based on original ViT model [44].

CROSS VALIDATION

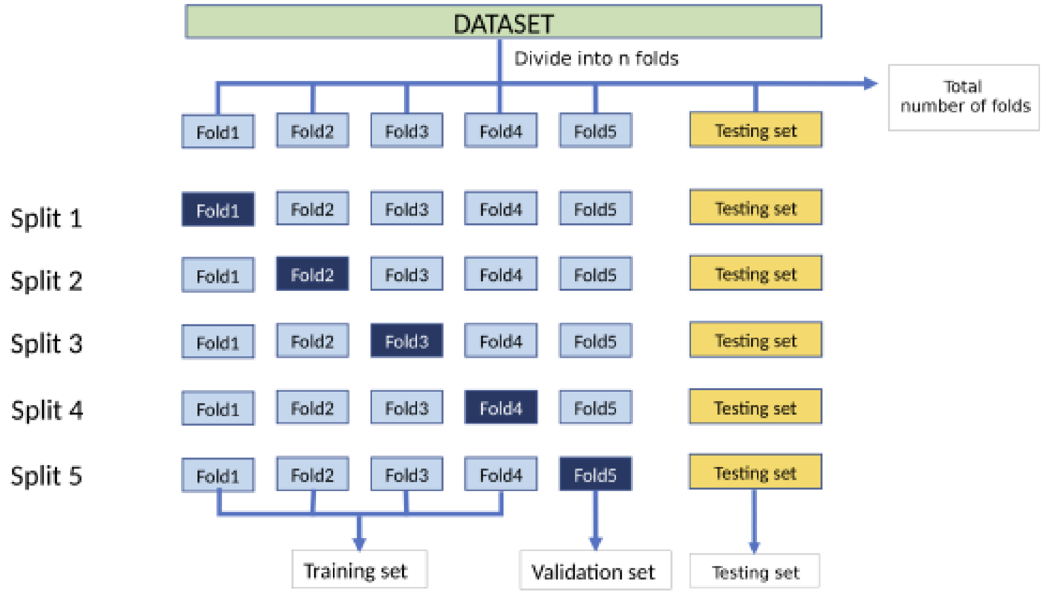


Figure 9. Stratified five-folds cross-validation, leaving one out for validation and the remaining 4 folds are used for training. Dark blue representing validation folds, light blue colour folds are used as training set and yellow colour folds are used as testing set containing unprocessed images. This generates 5 trained models.

CROSS VALIDATION

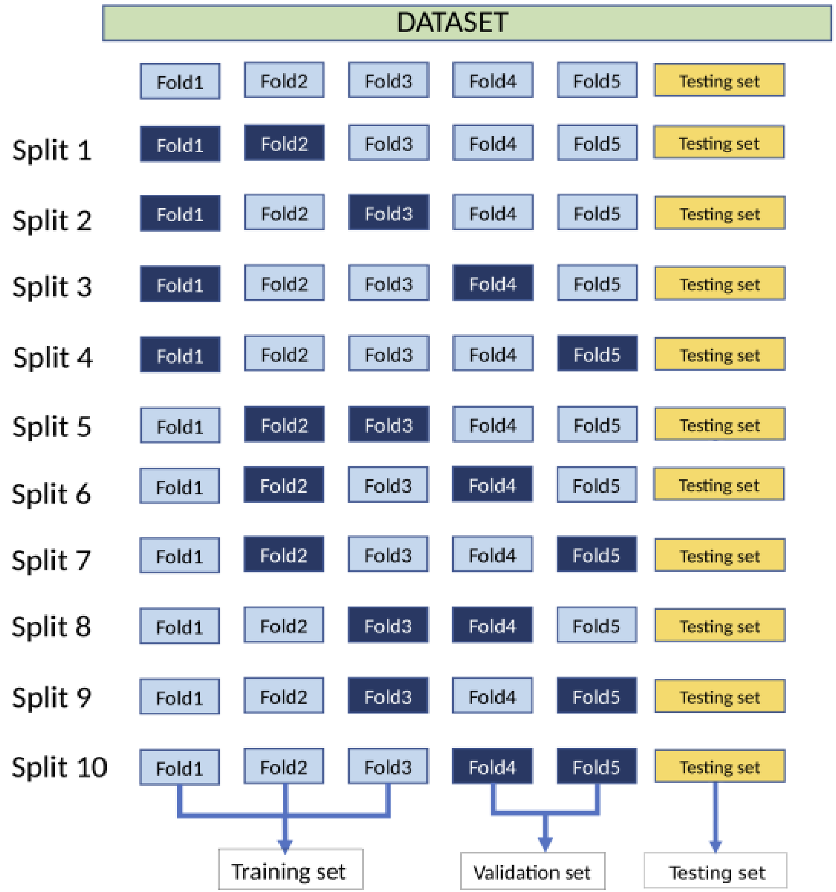


Figure 10. Stratified five-folds cross-validation and leaving two out as validation set and the rest are used for training resulting in 10 different models. Dark blue representing validation folds and light blue colour folds are used as training set.

CROSS VALIDATION

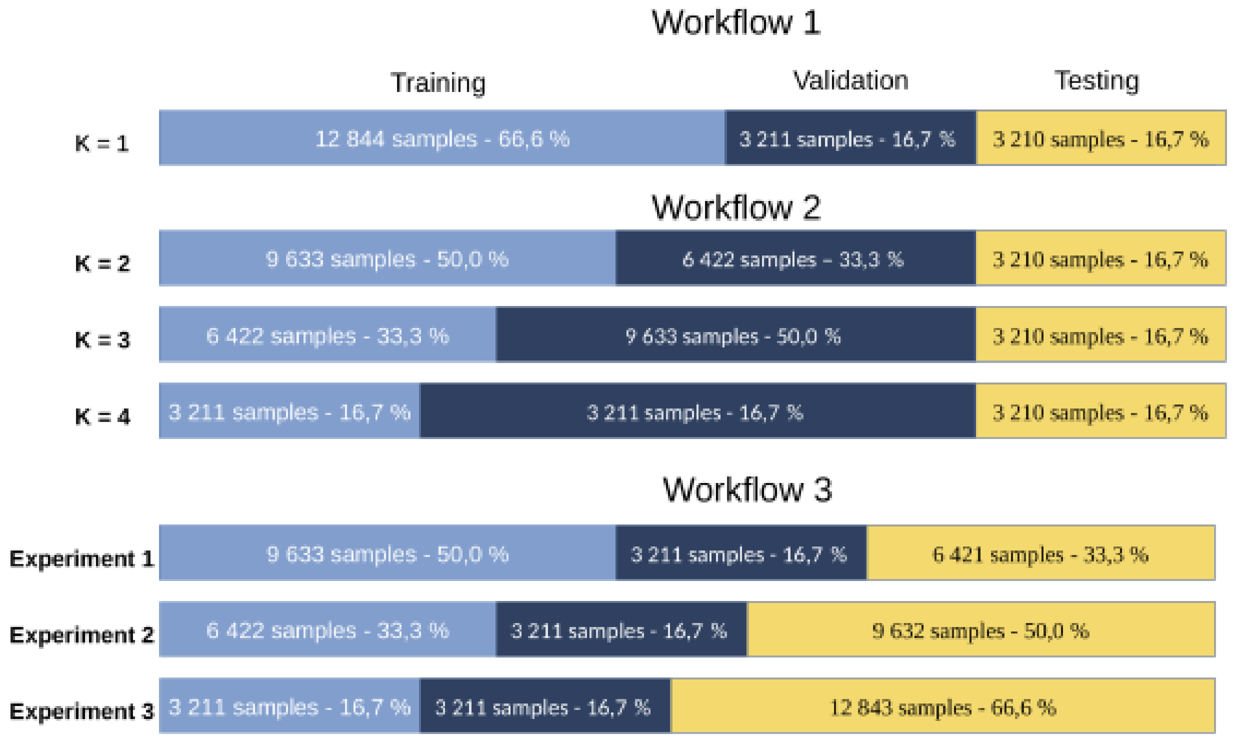


Figure 11. Variation of training/validation set and testing image set for the 3 workflows. The training/validation set is used for the cross-validation as shown in Figure 9.

EVALUATION METRICES

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)}$$

$$\mu_{F1-Score} = \frac{\sum_{i=1}^{\mathcal{N}} (F1 - Score_i)}{\mathcal{N}}$$
$$\sigma_{F1-Score} = \sqrt{\frac{\sum_{i=1}^{\mathcal{N}} (F1 - Score_i - \mu_{F1-Score})^2}{\mathcal{N}}}$$

RESULTS

Table 2. Comparison between state-of-the-art CNN-based models and vision transformer models on agricultural image classification. The F1-Score has been calculated using Equation (10) with $\mathcal{N} = 5$.

Model	$\mu_{F1-Score}$	μ_{Loss}
ViT B-16	**0.994 ± 0.002**	**0.656**
ViT B-32	**0.992 ± 0.002**	**0.672**
EfficientNet B0	0.987 ± 0.005	0.735
EfficientNet B1	0.989 ± 0.005	0.720
ResNet 50	0.992 ± 0.005	0.716

Although all network families obtain high accuracy and F1-Score, the classification of crops and weed images using vision transformer yields the best prediction performance.

REFERENCES

- Reedha, R.; Dericquebourg, E.; Canals, R.; Hafiane, A.Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images. Remote Sens. 2022, 14, 592. https://doi.org/10.3390/rs14030592

- https://viso.ai/deep-learning/vision-transformer-vit/


Thank you