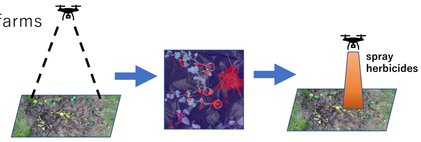
Deep Learning Neural Networks Based Weed Detection Using Aerial RGB And NIR Image Processing Methods For Smart Agriculture

JIAQI ZOU
Advanced Network Architecture Research Laboratory
Graduate School of Information Science and Technology
Osaka University

2023/8/21

Weed Detection Using Drone For Smart Agriculture

- significance of research
  - Increasing crop yields
  - Reducing the indiscriminate use of pesticides
  - Saving labor to manage farms



- Weeding process
  1. Drone captures farm images at high altitude
  2. Input the image into the neural network and find the positions
  3. Spray this area with the corresponding herbicide

Challenges in Implementing

1. Distinguish between weeds and crops
2. Aerial images of weeds are blurred
3. Weed labeling is hard

Proposal Approach

1. Use near-infrared images
2. Super-resolution
3. Self-supervised learning (contrastive learning)

Proposal Approach —— Near-Infrared （NIR） Images

NIR Image

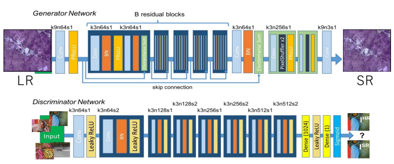
- Wavelength range of 700 to 1400 nanometers.
- Determine the health of plants
- The reflectance of NIR for different plants is different
- Normalized Difference Vegetation Index (NDVI)
  - $NDVI = \frac{NIR - Red}{NIR + Red}$
- This research using Red, Green, NIR 3-channels (RGN) images
- Problems:

  Not enough RGN images

Proposal Approach —— Super Resolution

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network (SRGAN)

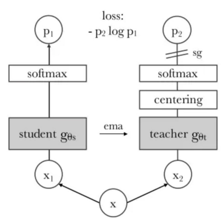
1. Low-resolution images (LR) are obtained by down-sampling high-resolution images (HR)
2. Input LR into the generator network to get super resolution images (SR)
3. Training the discriminator to distinguish between SR and HR
4. Training generator generates SR to confuses the discriminator



Proposal Approach —— Contrastive Learning

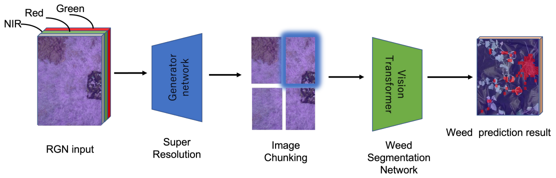
Emerging Properties in Self-Supervised Vision Transformers

- Same category should be in the same position in the feature space
- The student networks (g_{θ_s}) and teacher network (g_{θ_t}) are two Vits, turn input image into a vector by self attention
- x1 input into g_{θ_s} is the local picture, x2 input into g_{θ_t} is the global picture
- Preventing Collapse
  1. Momentum update
  2. Centering
- Deeper MLP



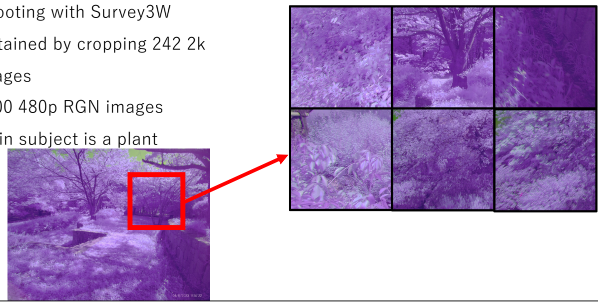
Proposed Architecture

1. Input RGN images taken by drones
2. Improvement of image resolution by generators network in SRGAN
3. Split the super-resolution image into pieces of 480x480
4. Weed Segmentation network is student network obtained by DINO training
5. The segmentation of weeds is obtained based on the self-attention of Vision Transformer



Dataset For Super Resolution

- Shooting with Survey3W
- Obtained by cropping 242 2k images
- 1500 480p RGN images
- main subject is a plant



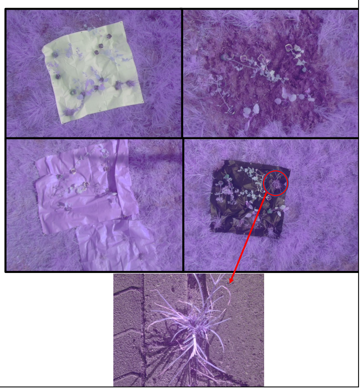
Dataset For Weeds Segmentation

4 types of RGN images taken by drones
403 images

- 33 Weeds and plants in white sheet
- 80 Weeds and plants in blue sheet
- 33 Weeds and plants in black sheet
- 227 Weeds and plants in real environment

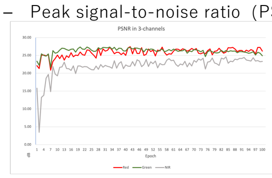
Cropped out the weeds and paste them to different backgrounds

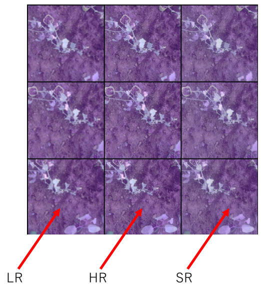
- 400 480p weeds in the center images



Experimental Results and Conclusions 1

1. Pre-training on ImageNet
2. Fine-tuning on local datasets
   - Pre-training is very effective
   - Red and green channels converge faster than NIR channels
   - Peak signal-to-noise ratio（PSNR）



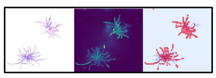


Experimental Results and Conclusions 2

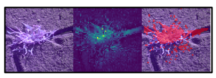
1. Pre-training on ImageNet
2. Fine-tuning on local datasets
   - Two scales of ViT were trained
   - Testing on three different types of test sets
   - Comparison of before and after fine-tuning using local datasets

Models / Image Type	Vit-S/16 (proposal)	Vit-S/8 (pre-train)	Vit-S/8 (proposal)
Simple background	81.4	80.7	83.3
complex context	74.4	69.3	77.8
real environment	40.1	27.3	44.2

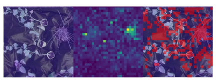
Experimental Results and Conclusions 3



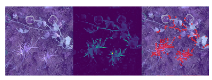
Performance of ViT-S/8 after fine-tuning on simple background



Performance of ViT-S/8 after fine-tuning on complex context



Performance of ViT-S/8 before fine-tuning on real environment

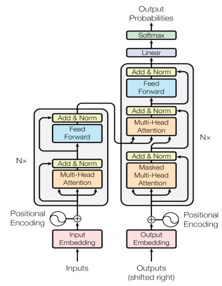


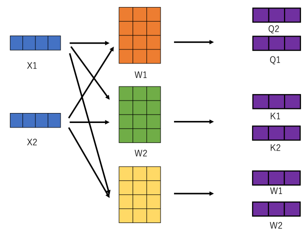
Performance of ViT-S/8 before fine-tuning on real environment

Look Forward To The Future

- Larger scale neural networks
  - ➢The larger the model of the transformer, the higher the vector dimension and the more accurate the features.
- Creating larger scale dataset
  - ➢transformer requires more training data to train self-attention due to lack of inductive bias
- Exploring better self-supervised learning models

Appendix 1





Appendix 2