# Few-shot Plant Disease Classification

Siddharth Khandelwal
190070062

# Outline

- Motivation
- Literature Survey
- Datasets
- Few-shot learning
- CLIP
- Prompt Learning
- Implementation
- Experiments
- Conclusion
- References

# Motivation

In a real world scenario one might not have many examples of classes of plant diseases on fields due to various factors :

1. Less documentation
2. Rare diseases

Few shot learning might prove helpful for such cases, to model and classify such few sampled diseases

# Literature Survey

Bacterial Wilt Disease Detection

- Attacks more than 200 species of crops
- BWENet for grading severity of bacterial wilt disease
- Multi-spectral and Multi fractal analysis

Multi-modal approaches

- Image-text modality
- Hyperspectral imaging
- Soil moisture, temperature, rainfall

# Datasets

## Plant Village Dataset

- 256 x 256 size images
- 38 Plant leaf classes
- 54000 total images

Corn Northern Leaf Blight

Grape Leaf blight

Peach Bacterial_spot

Source : (https://data.mendeley.com/datasets/tywbtsjrjv/1)

## Plant Doc Dataset

- Varying image sizes
- 28 Plant leaf classes
- 2600 total images

Bell pepper Leaf Spot

Tomato Early Blight

Healthy Apple leaf

Source : (https://github.com/pratikkayal/PlantDoc-Dataset)

# Few-shot learning

Classify new data when you have only a few training samples with supervised information

**N-way-K-shot classification :**

- **Training set :**
  Consists of N class labels each with K labelled images (where K can be around 5 to 10)
- **Query set :**
  Contains the query images which need to be classified among the N classes

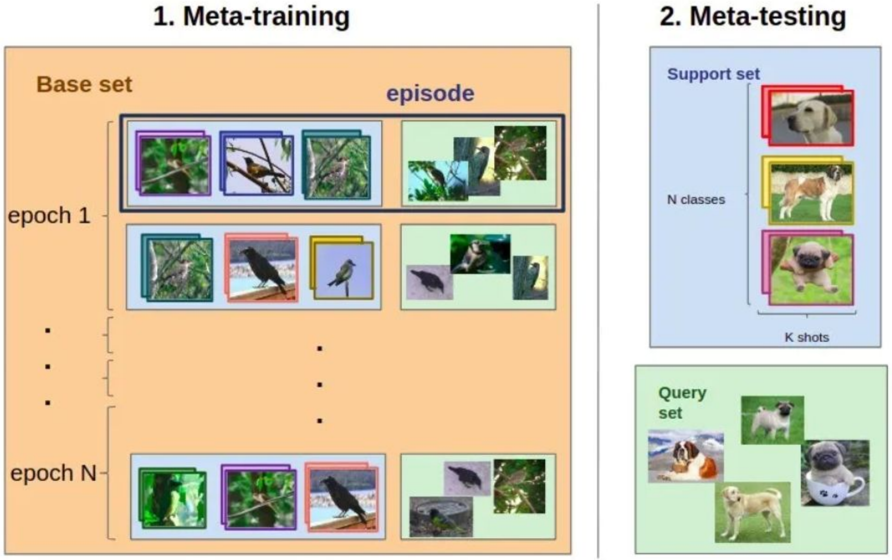
Possible approach : Gain experience from other similar problems. To this end, most approaches characterize few-shot learning as a **meta-learning** problem.

# Meta Learning

We treat our N-way-K-shot classification problem as the TEST data and a similar large base dataset as the meta learning training set (TRAIN)

**Training:**
N classes and K support images per each classes along with Q query images from the large TRAIN set

**Test:**
Classify our less labelled dataset using this trained model in similar episodes.

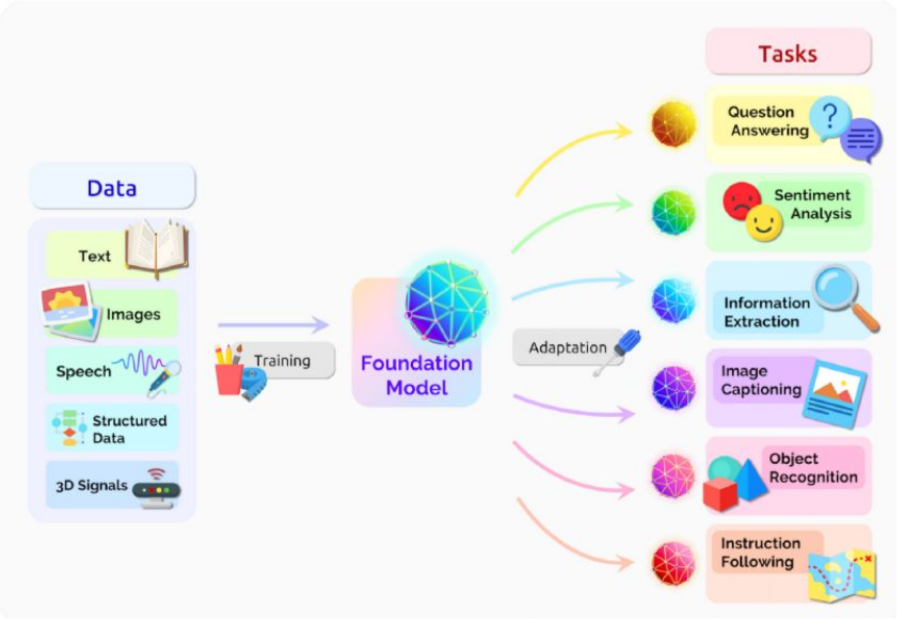



Meta learning [1]


[1] E. Bennequin, “Few-Shot Image Classification with Meta-learning”,2018

# Foundation Models in AI

- Models that are trained on a broad set of unlabeled data
- Can be adapted for different tasks, with minimal fine-tuning
- Various application using self-supervised learning and transfer learning
- Eg: BERT, GPT-3, DALL-E 2, CLIP etc

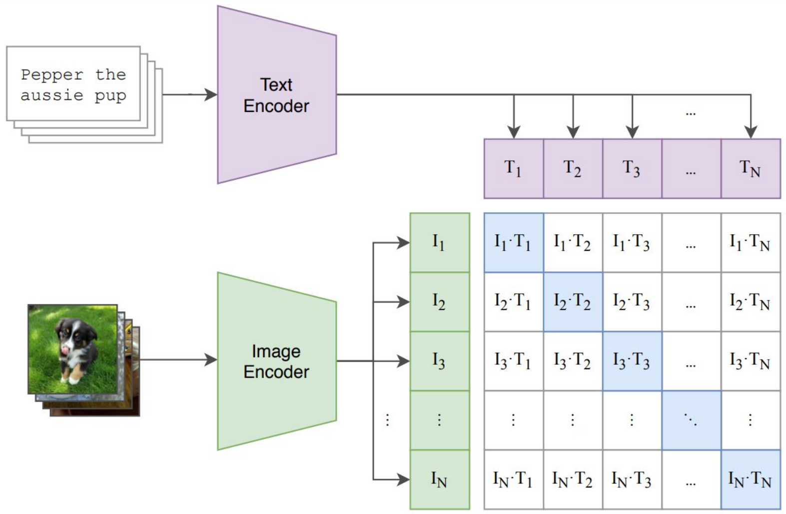



Foundation Models in AI [2]

[2] Bommasani, R., Hudson, D., Adeli, E., Altman, R., & Arora et. al. (2021). “On the Opportunities and Risks of Foundation Models”.

# CLIP (Contrastive Language-Image Pre-training)

- Trained on wide variety of images and corresponding text from the internet
- During pre-training the model takes in a batch of N pairs of (image,text) input and learns a multi-modal embedding space by jointly training an image encoder and a text encoder
- Maximize the cosine similarity of the true N pairs and minimizing those of the rest possible pairs




Clip Model Pretraining [3]

[3] Radford, A., Kim, J., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I.. (2021). Learning Transferable Visual Models From Natural Language Supervision.

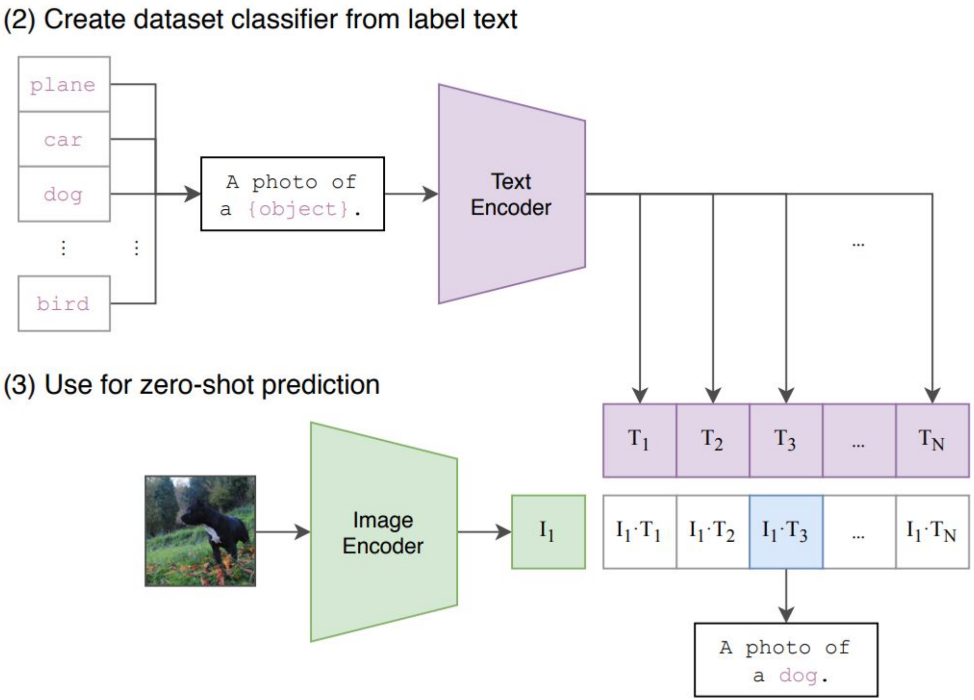
# Zero-shot with CLIP

Aim : To classify an unseen image into infected or healthy using CLIP without training on Plant Village

Method : Visual classification by providing the names of the visual categories to be recognized to the pre-trained CLIP

Divided Plant Village Dataset into healthy and infected leaf classes

Result : **85 %** accuracy




Zero-shot with CLIP [3]

# Problems with Zero shot using CLIP

- Prompt engineering is difficult
- Small changes in prompt result in large accuracy changes
- Cannot classify diseases

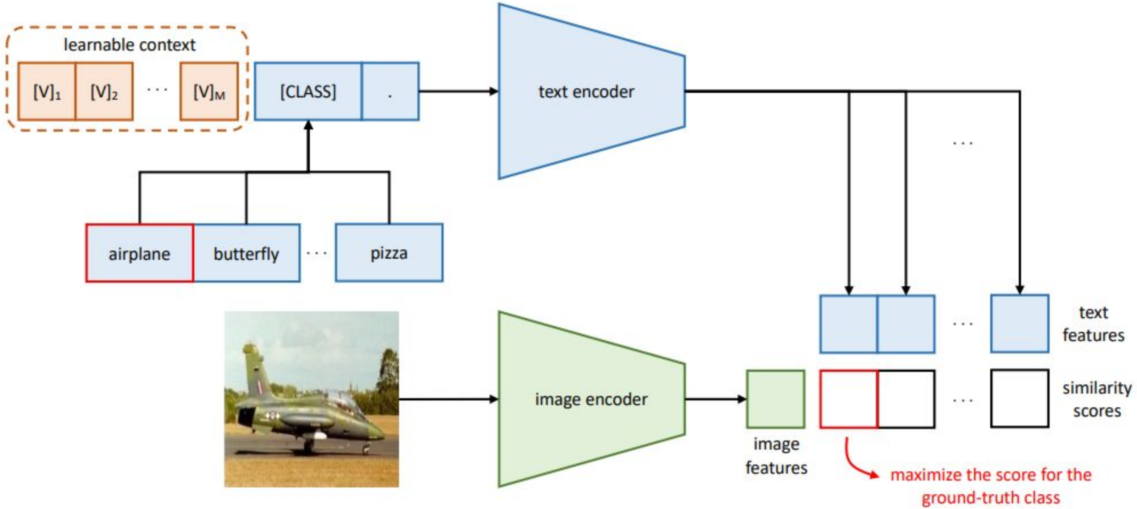
Need an automatic way to decide prompts for classes for better classification accuracy

- CoOp
- CoCoOp

# Prompt Learning

## A) **Context Optimization (CoOp Model)**

- Models a prompt's context words with learnable vectors
- Learning context through back propagation of classification loss
- Context learnt is static



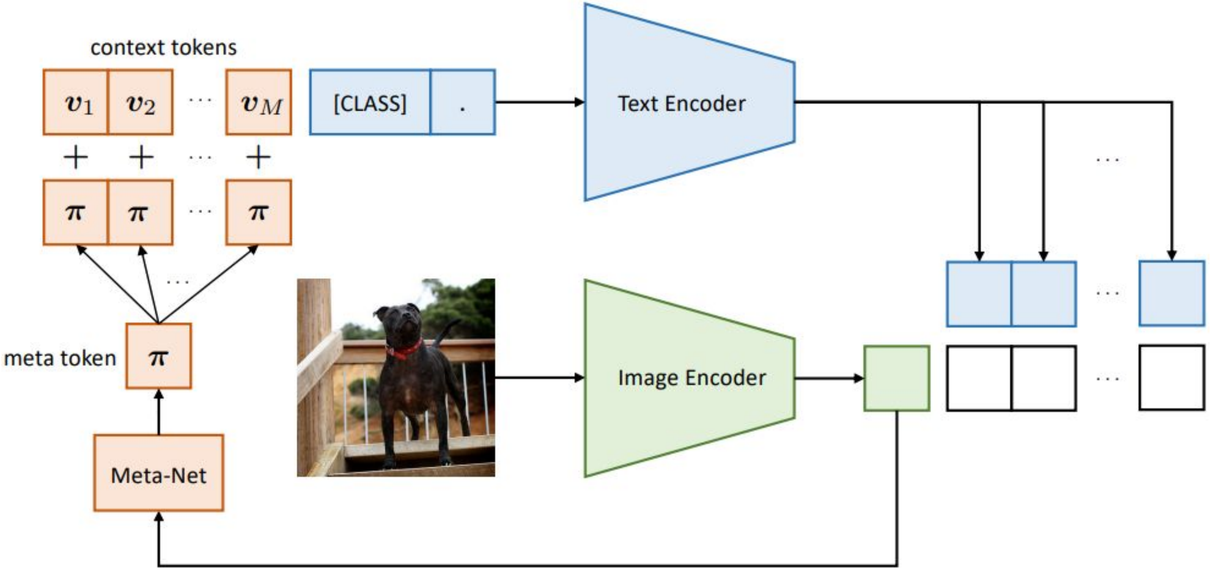

CoOp Model [4]

[4] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, & Ziwei Liu (2022). Learning to Prompt for Vision-Language Models. *International Journal of Computer Vision, 130(9), 2337–2348.*

# B) Conditional Context Optimization (CoCoOp)

- Dynamic prompts
- Neural network on each instance (image) which adds information to context
- Better generalizability on unseen classes than CoOp

CoCoOp Model [5]





[5] Zhou, K., Yang, J., Loy, C., & Liu, Z.. (2022). Conditional Prompt Learning for Vision-Language Models

# Implementation

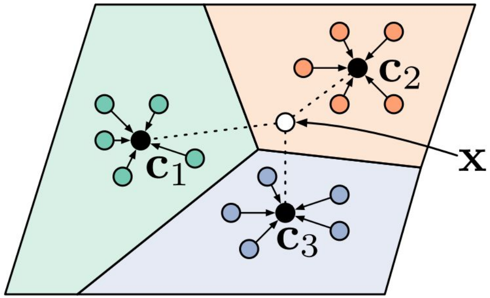
Python framework : Pytorch

System : NVIDIA GeForce GTX 1080 Ti of VIP Lab

# Experiments

Prototypical Few Shot Classification on Plant Doc Dataset

We see a drastic decrease in accuracy due to more complex images / noisy backgrounds etc in Plant Doc



(a) Few-shot

| | Plant Village Dataset | | Plant Doc Dataset | |
|---|---|---|---|---|
| | Training Accuracy (%) | Test Accuracy (%) | Training Accuracy (%) | Test Accuracy (%) |
| 5-way-5-shot | 93 | 82.8 | 82 | 46 |
| 5-way-10-shot | 96 | 87 | 85 | 48 |

# Experiments Contd.

**Linear Probing on CLIP**

- Logistic regression of features obtained from CLIP’s image encoder
- Dataset : Plant Village
- Training & Validation : 38 classes having N labelled examples each (N shot)
- Test : 38 classes with around 300 examples each

| | 1 shot | 2 shot | 4 shot | 8 shot | 16 shot |
|---|---|---|---|---|---|
| Accuracy (%) | 58.8 % | 76.3 % | 87 % | 91.4 % | 94.2 % |

Note: The test set and training set have common classes between them in this case

# Experiments Contd.

**Few shot learning using CoOp on Plant Village**

- Dataset : Plant Village
- Training & Validation : 38 classes having K labelled examples each
- Test : 38 classes with around 300 examples each

| | 1 shot | 2 shot | 4 shot | 8 shot | 16 shot |
|---|---|---|---|---|---|
| Accuracy (%) | 27 ± 3 | 50 | 53 ± 5 | 74 ± 2 | 82 |

Note: The test set and training set have common classes between them in this case

# Experiments Contd.

**Generalization on Unseen New Classes**

- CoOp and CoCoOp are compared
- Dataset : Plant Village
- Training & Validation set : 30 classes with around 1000 examples each
- Test set : 8 new classes with 20 examples each

| | CoOp | | CoCoOp | |
|---|---|---|---|---|
| | Training Accuracy (%) | Test Accuracy (%) | Training Accuracy (%) | Test Accuracy (%) |
| 5-shot | 78 | 26 | 58 | 38 |
| 10-shot | 86 | 13 | 66 | 20 |

# Conclusion

- Learning with limited examples per class with Linear Probe and using CoOp feasible
- CoCoOp and CoOp models does not give good results on Plant Village for generalization on unseen classes

# References

[1] E. Bennequin, “Few-Shot Image Classification with Meta-learning”,2018

[2] Bommasani, R., Hudson, D., Adeli, E., Altman, R., & Arora et. al. (2021). “On the Opportunities and Risks of Foundation Models”

[3] Radford, A., Kim, J., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I.. (2021). Learning Transferable Visual Models From Natural Language Supervision.

[4] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, & Ziwei Liu (2022). Learning to Prompt for Vision-Language Models. *International Journal of Computer Vision, 130(9), 2337–2348.*

[5] Zhou, K., Yang, J., Loy, C., & Liu, Z.. (2022). Conditional Prompt Learning for Vision-Language Models

Thank You