JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Efficient and Intelligent Computing Lab

Hint-Aug: Drawing Hints From Foundation Vision Transformers Towards Boosted Few-Shot Parameter-Efficient Tuning

Zhongzhi Yu, Shang Wu, Yonggan Fu, Shunyao Zhang, and Yingyan (Celine) Lin

Georgia Institute of Technology

Georgia Tech

Paper Tag: WED-AM-273

Challenge: Scarcity of Data

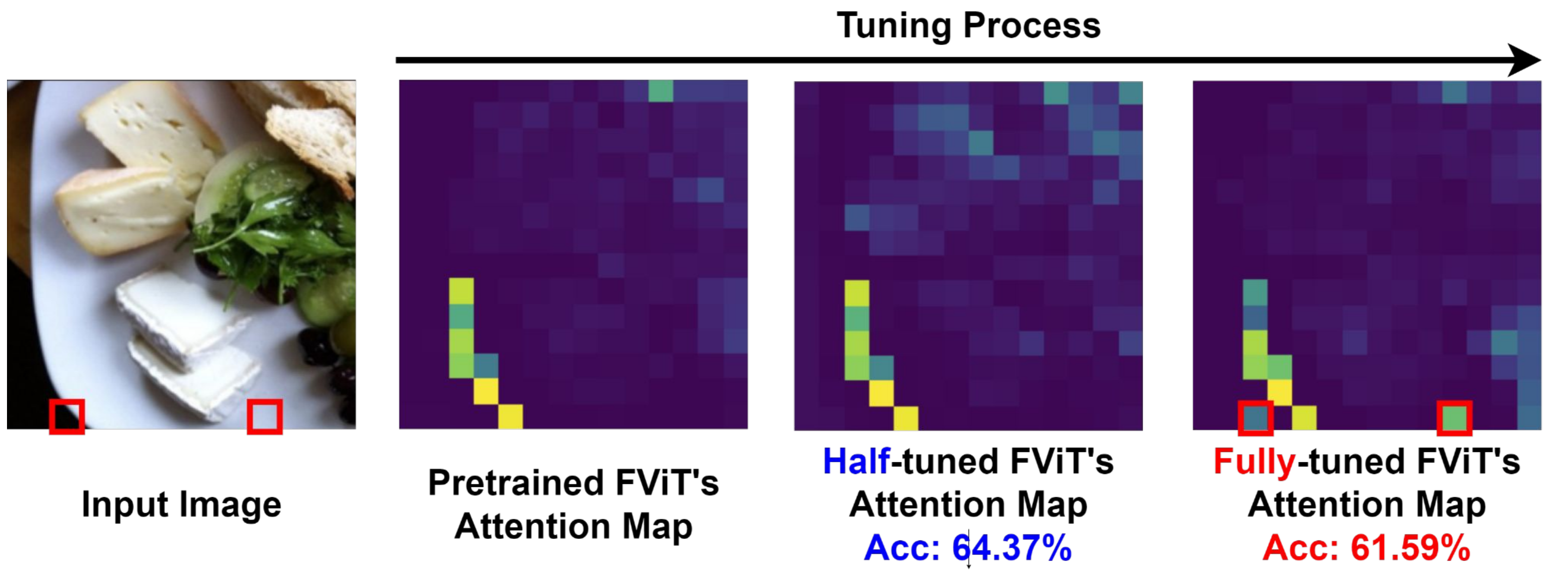
- Despite PET's great promise, collecting sufficient downstream data is arduous
- **Few-shot tuning**: A common scenario
  - Tuning with limited samples per class
  - Largely impacts PET performance
    - 1-shot **70.2%** vs. 1000-shot **90.4%** @Pet dataset *[Zhang, arXiv'22]*

How to make better use of few-shot tuning data?

- An effective augmentation pipeline
  - Where to augment?
  - How to augment?

Insight: Shift in Attention Indicates Over-fitting

- During PET, FViT's attention shifts to **irrelevant** positions (**red boxes**)



➡ Shift in attention map indicates potential over-fitting

Leverage the pretrained FViT to guide the augmentation of few-shot PET

Hint-Aug: Key Enablers

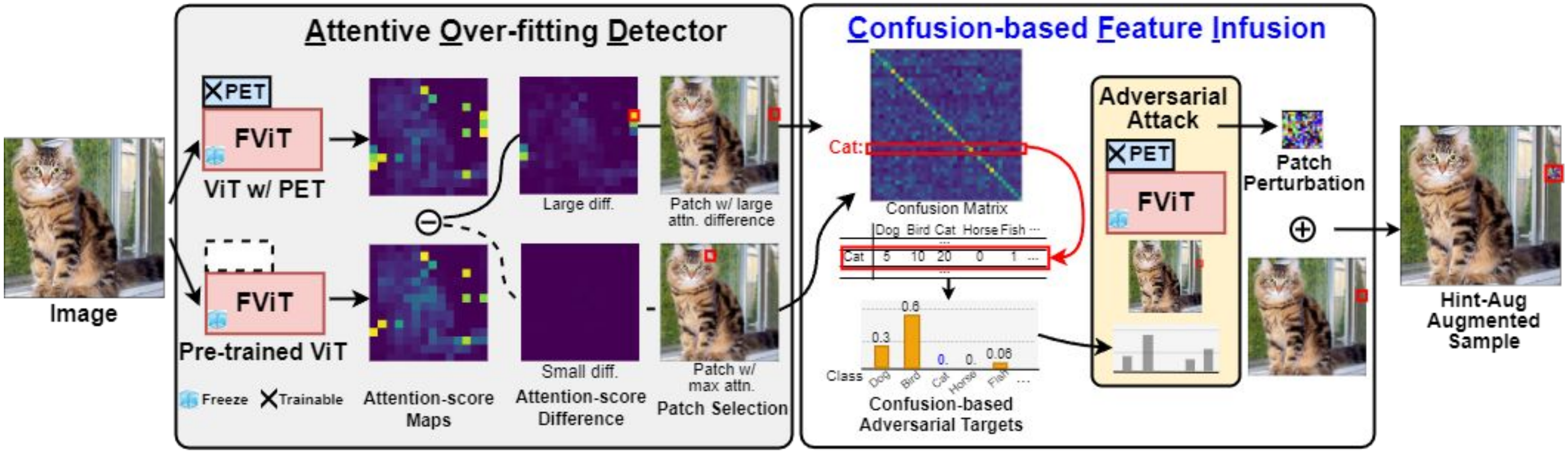
- Core Idea: Leverage the pretrained FViT's learned **generalizable features** to guide augmentation

Q1: **Where to augment?** ➡ ***A1***: **Attentive Over-fitting Detector**

Augment the patch that the FViT is over-fitted to

Q2: **How to augment?** ➡ ***A2***: **Confusion-based Feature Infusion**

Infuse easy-to-confuse features from FViT



Hint-Aug: Drawing Hints From Foundation Vision Transformers Towards Boosted Few-Shot Parameter-Efficient Tuning

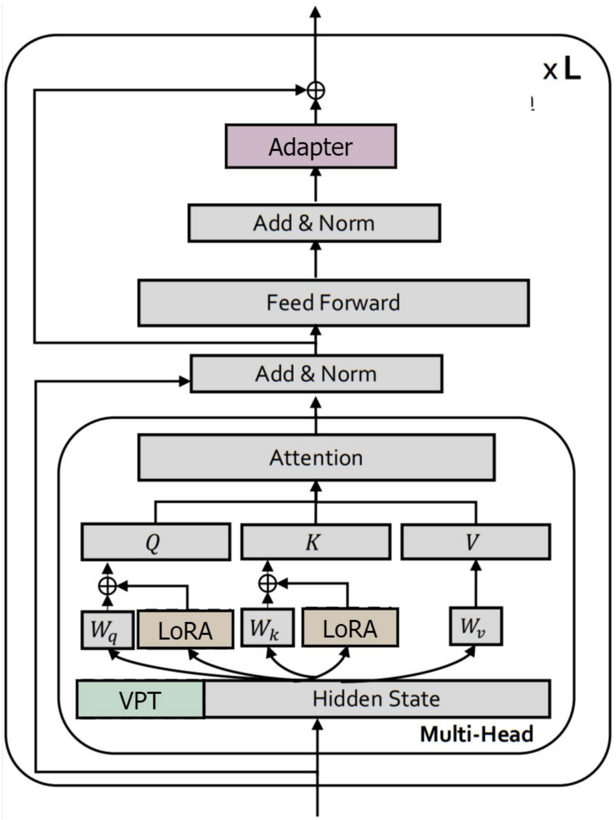
Zhongzhi Yu, Shang Wu, Yonggan Fu, Shunyao Zhang, and Yingyan (Celine) Lin

Georgia Institute of Technology

Paper Tag: WED-AM-273

Background: Parameter-efficient Tuning (PET)

- Foundation vision transformers (FViTs) learns features w/ strong **adaptation** ability

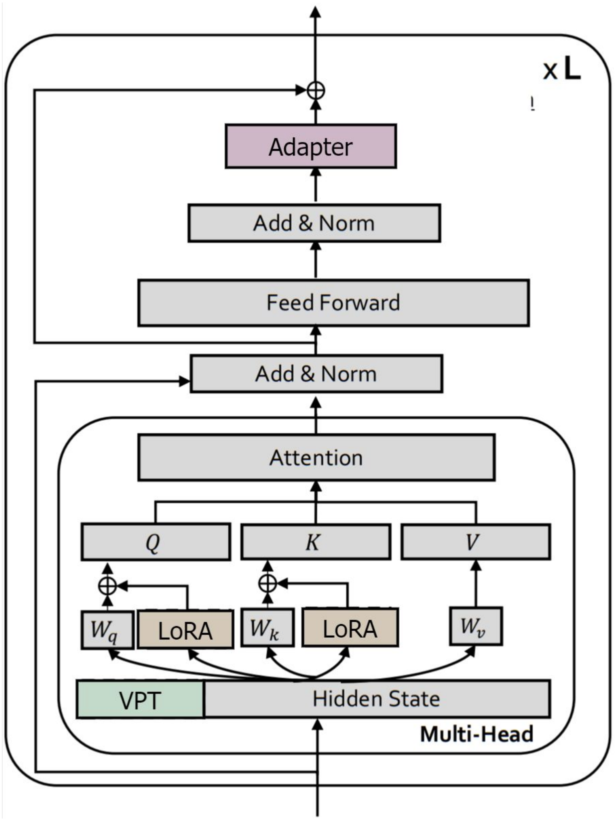


[Zhang, arXiv'22]

Background: Parameter-efficient Tuning (PET)

- Foundation vision transformers (FViTs) learns features w/ strong **adaptation** ability

- This motivates **PET**: tune FViTs with **limited trainable params**



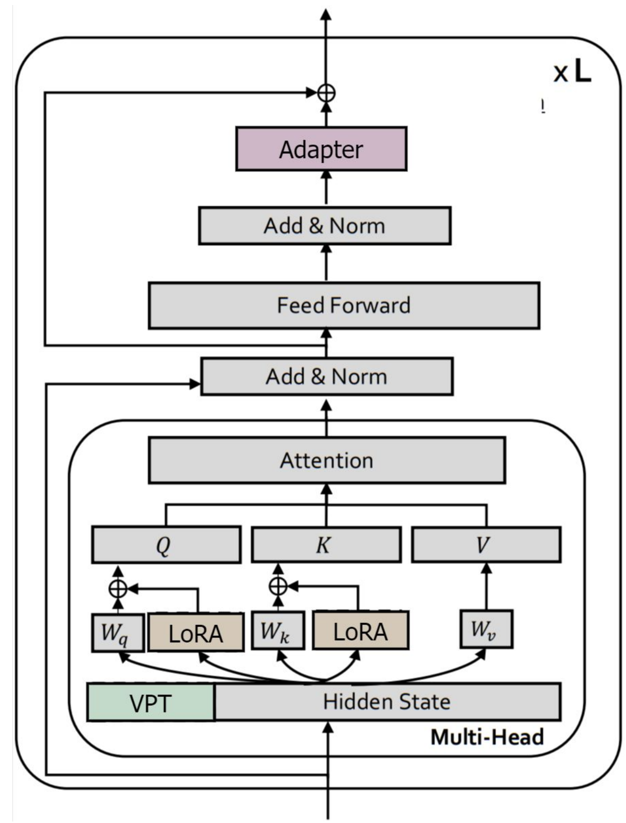
[Zhang, arXiv'22]

Background: Parameter-efficient Tuning (PET)

- Foundation vision transformers (FViTs) learns features w/ strong **adaptation** ability

- This motivates **PET**: tune FViTs with **limited trainable params**

- Compared with full tuning
  - Reduced storage cost
  - Promising accuracy



[Zhang, arXiv'22]

Challenge: Scarcity of Data

- Despite PET's great promise, collecting sufficient downstream data is arduous

Challenge: Scarcity of Data

- Despite PET's great promise, collecting sufficient downstream data is arduous

- **Few-shot tuning**: A common scenario
  - Tuning with limited samples per class
  - Largely impacts PET performance
    - 1-shot **70.2%** vs. 1000-shot **90.4%** @Pet dataset *[Zhang, arXiv'22]*

Our Goal: Improve Data Efficiency

How to make better use of few-shot tuning data?

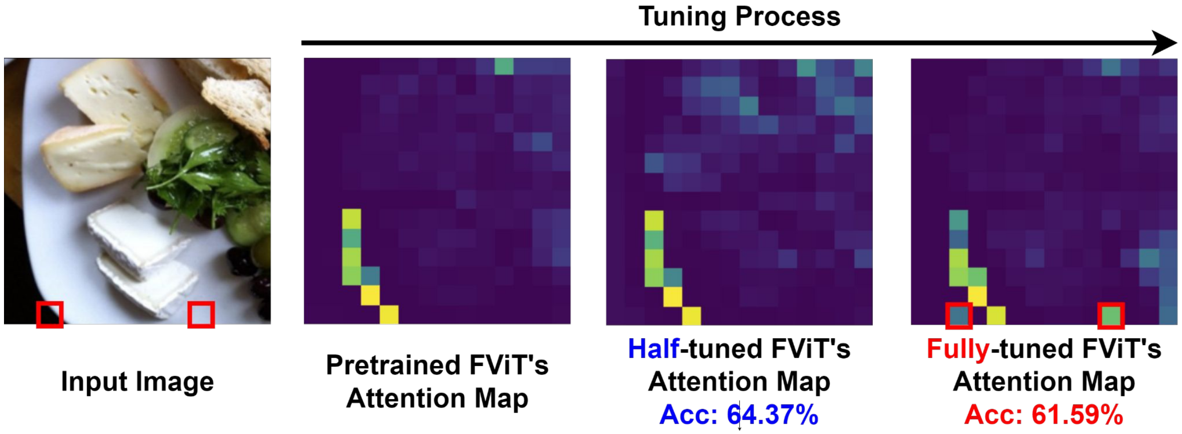
- An effective augmentation pipeline

  Where to augment?

  How to augment?

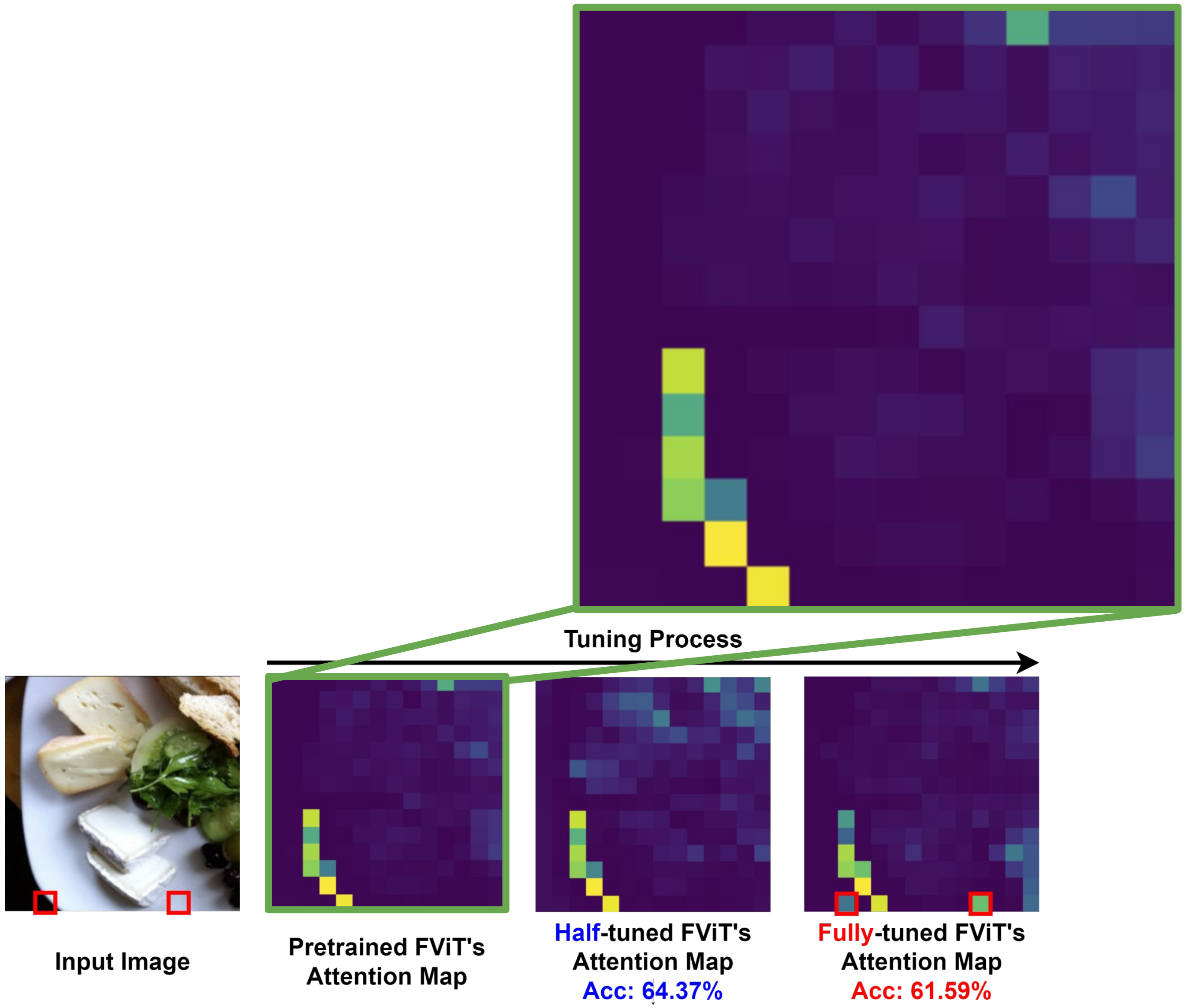
Insight: Shift in Attention Indicates Over-fitting

- During PET, FViT's attention shifts



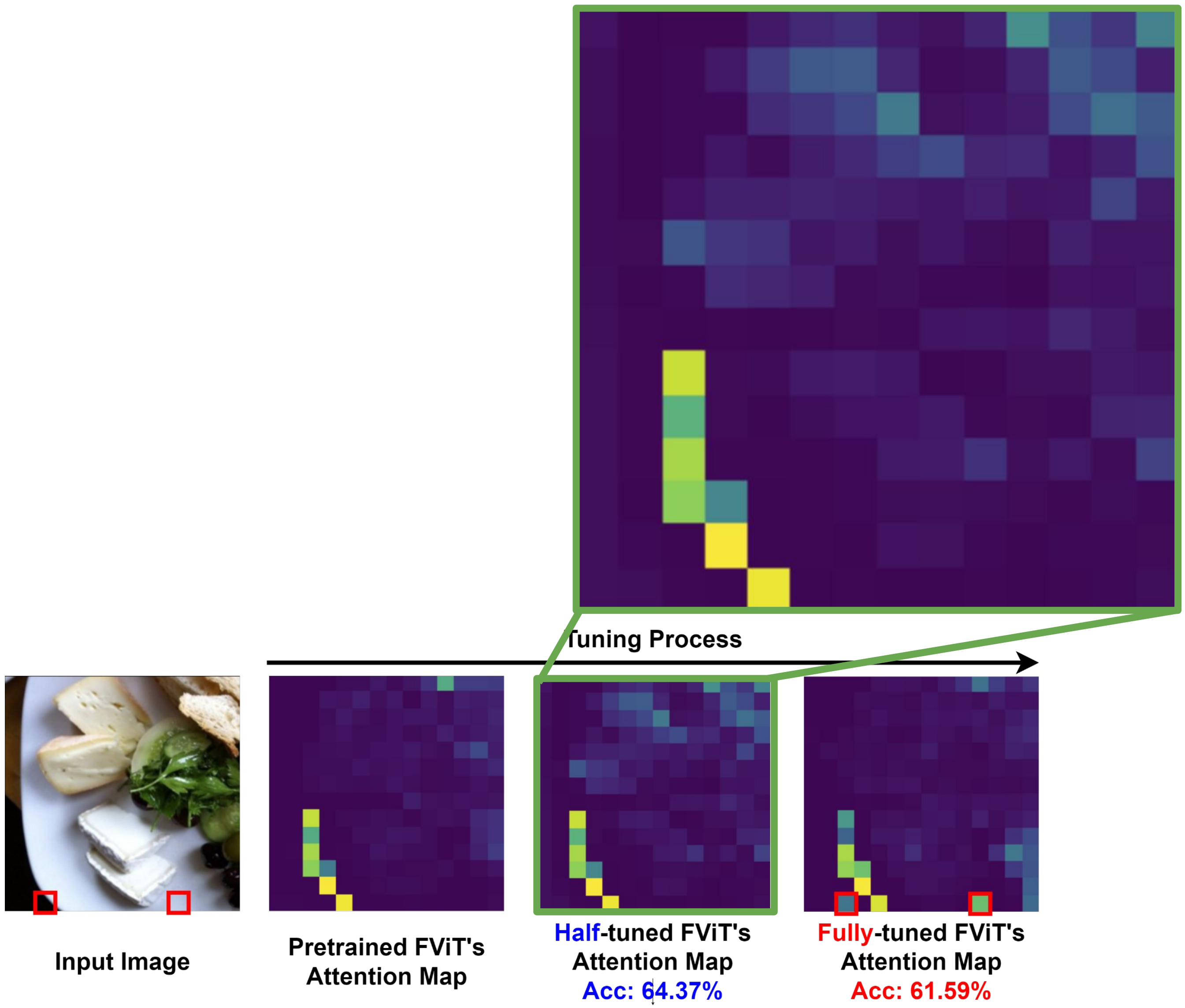
Insight: Shift in Attention Indicates Over-fitting

- During PET, FViT's attention shifts



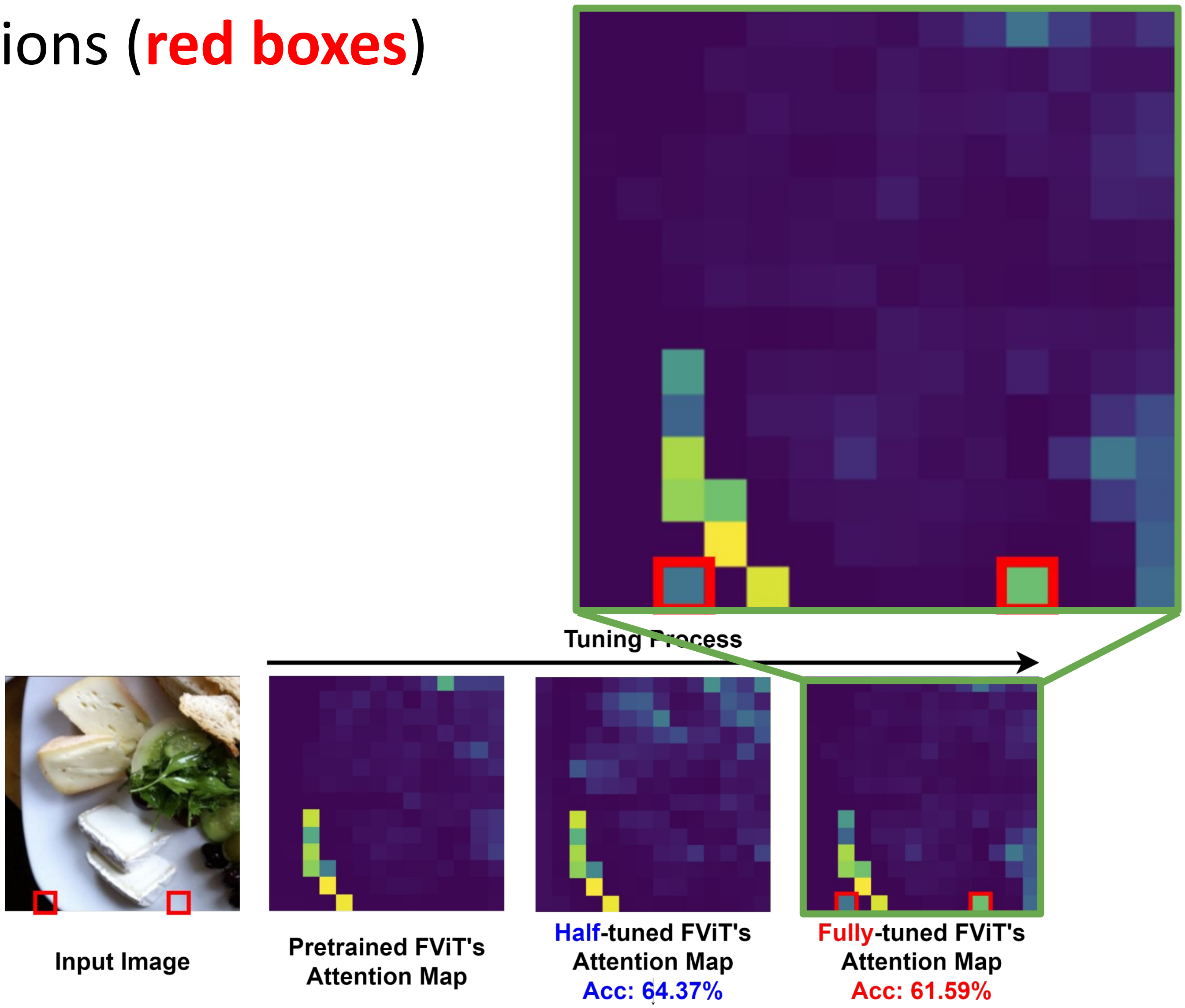
Insight: Shift in Attention Indicates Over-fitting

- During PET, FViT's attention shifts



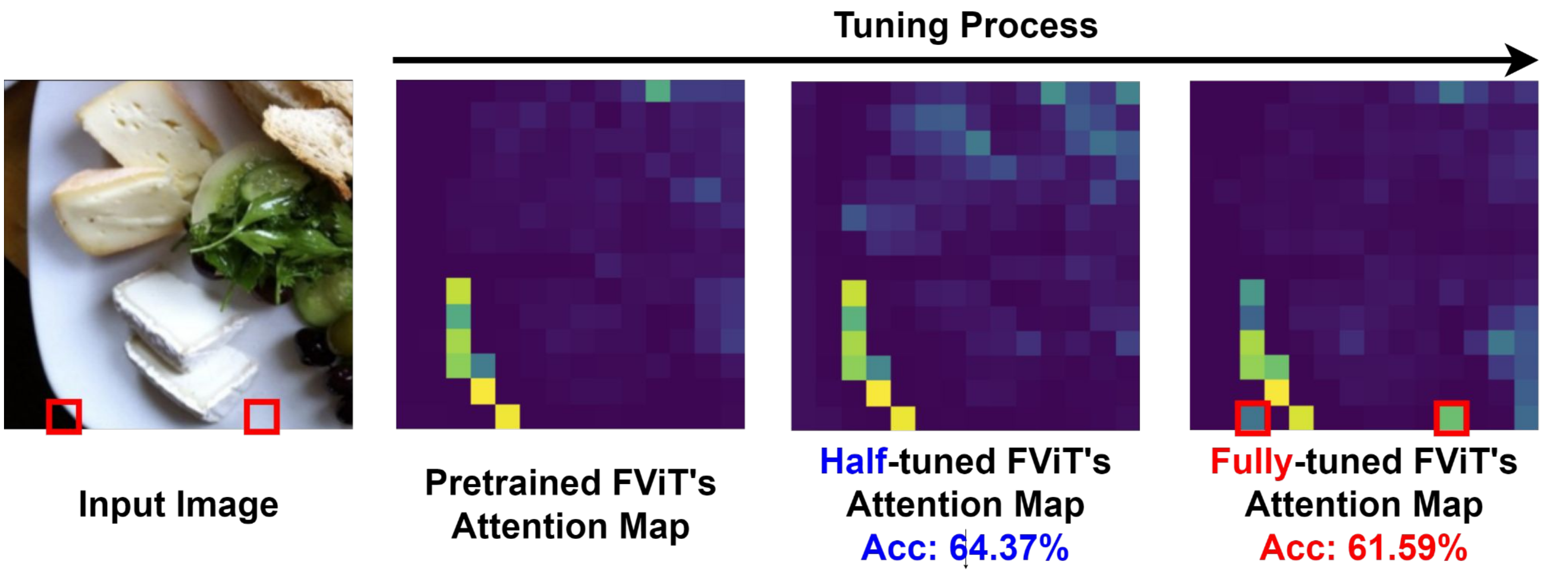
Insight: Shift in Attention Indicates Over-fitting

- During PET, FViT’s attention shifts to **irrelevant** positions (**red boxes**)



Insight: Shift in Attention Indicates Over-fitting

- During PET, FViT's attention shifts to **irrelevant** positions (**red boxes**)



➡ Shift in attention map indicates potential over-fitting

Leverage the pretrained FViT to guide the augmentation of few-shot PET

Our Contributions

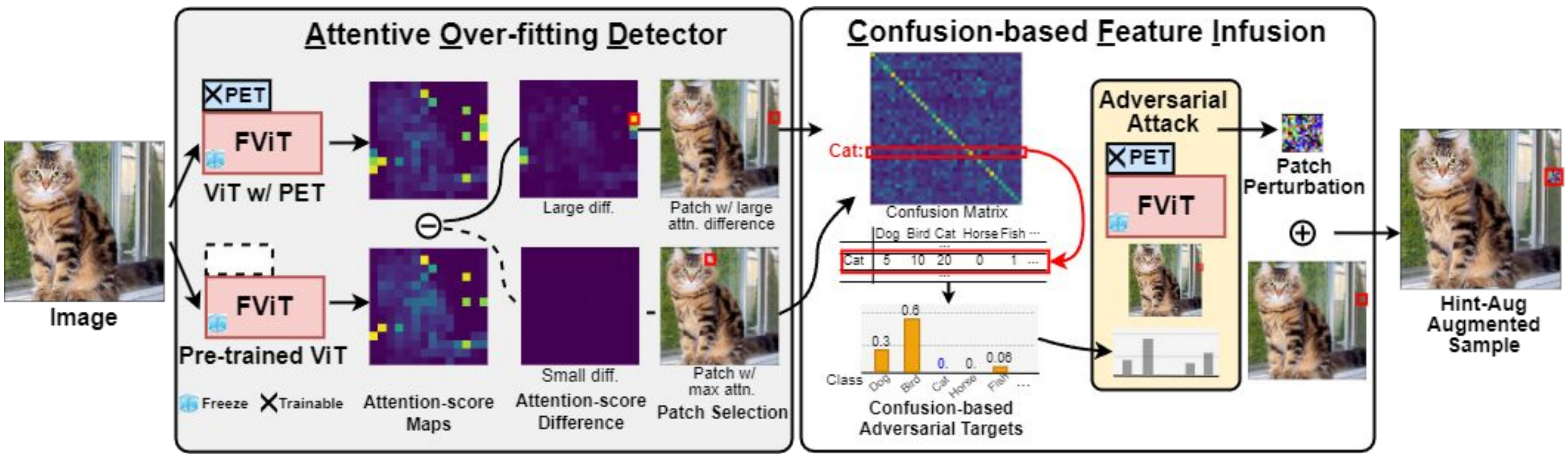
- Propose **Hint**-based Data **Aug**mentation (**Hint-Aug**) to guide data augmentation in few-shot PET

- Integrate two key enablers:
  - **Attentive Over-fitting Detector**: identify the over-fitting samples with attention maps
  - **Confusion-based Feature Infusion**: infuse pretrained FViTs' learned features to data

- SOTA accuracy-data efficiency trade-off: e.g., a **2.22%** higher accuracy with **50%** less data on Pet dataset

Hint-Aug: Core Idea

- Leverage the pretrained FViT's learned **generalizable features** to guide augmentation



Hint-Aug: Key Enablers

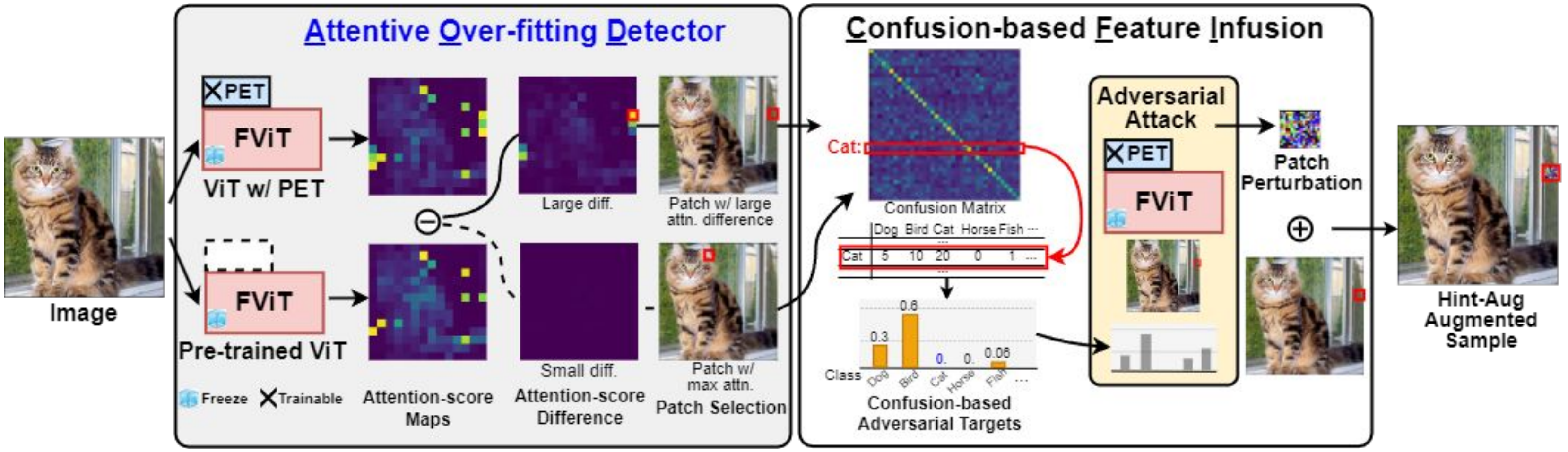
***Q1*: Where to augment? → *A1*: Attentive Over-fitting Detector**

Detect and augment the patch that FViT is over-fitted to

- Attention map diff. between pretrained and tuned FViT
  - Avg. diff **>** threshold: Suspicious to **over-fitting**

    → Select **largest diff.** patch
  - Avg. diff **<=** threshold: No significant over-fitting

    → Select **highest attention** patch



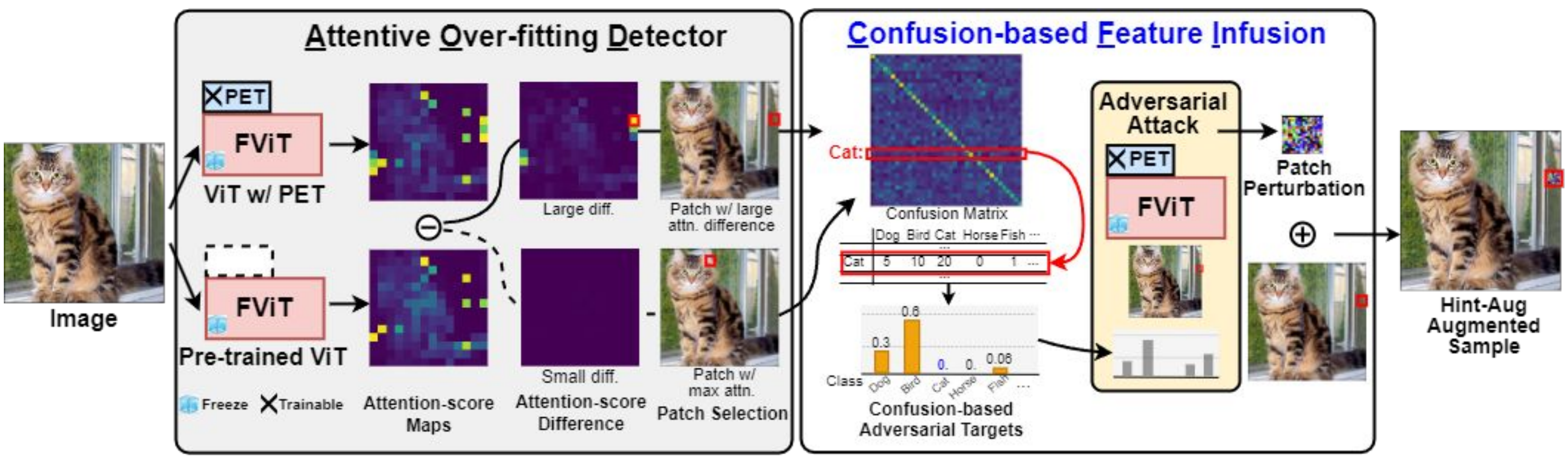
Hint-Aug: Key Enablers

Q1: **Where to augment?** ➡ *A1*: **Attentive Over-fitting Detector**

Q2: **How to augment?** ➡ *A2*: **Confusion-based Feature Infusion**

Infuse easy-to-confuse features to FViT

- Calculate **confusion-based adversarial targets** C based on prob. of wrongly classified to each class
- Infuse features to selected patch w/ adv. attack with target C

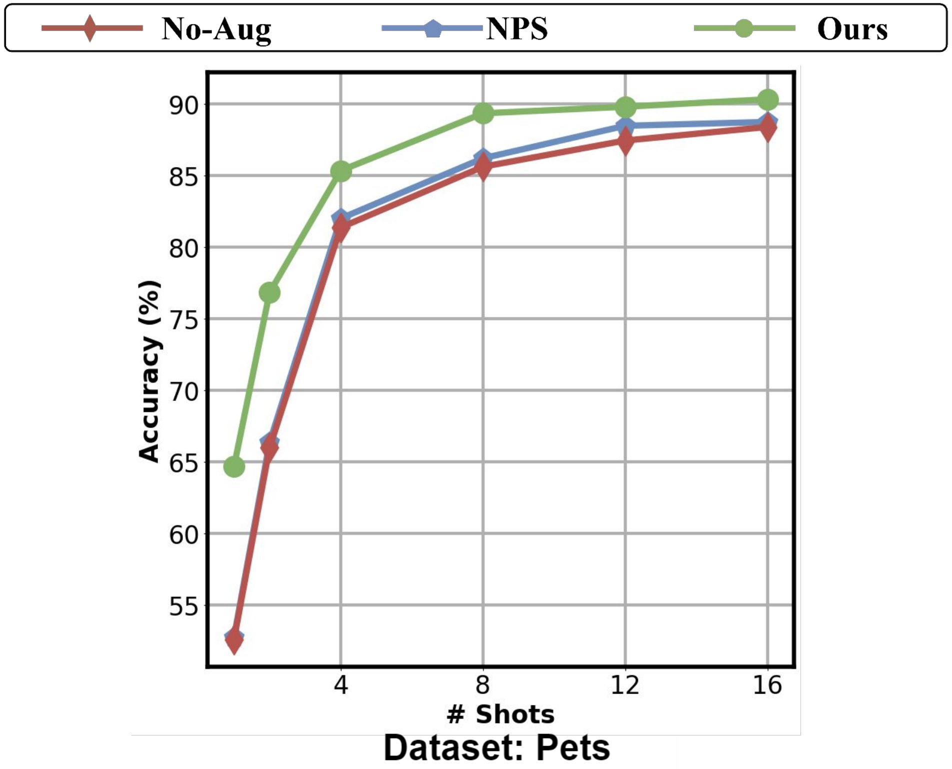


Hint-Aug: Evaluation Settings

- **Three** PET methods:
  - Adapter *[Houlsby, ICML'19]*, LoRA *[Hu, arXiv'21]*, VPT *[Jia, ECCV'22]*
- **Five** few-shot datasets:
  - Food, Pet, Flowers, Aircraft, Cars
- **Eight** few-shot settings: 1/2/4/8/12/16-shot
- FViT: ImageNet pretrained ViT-Base [Dosovitskiy, ICML'20]
- **Two** SOTA baselines: No augment; NPS *[Zhang, arXiv'22]*

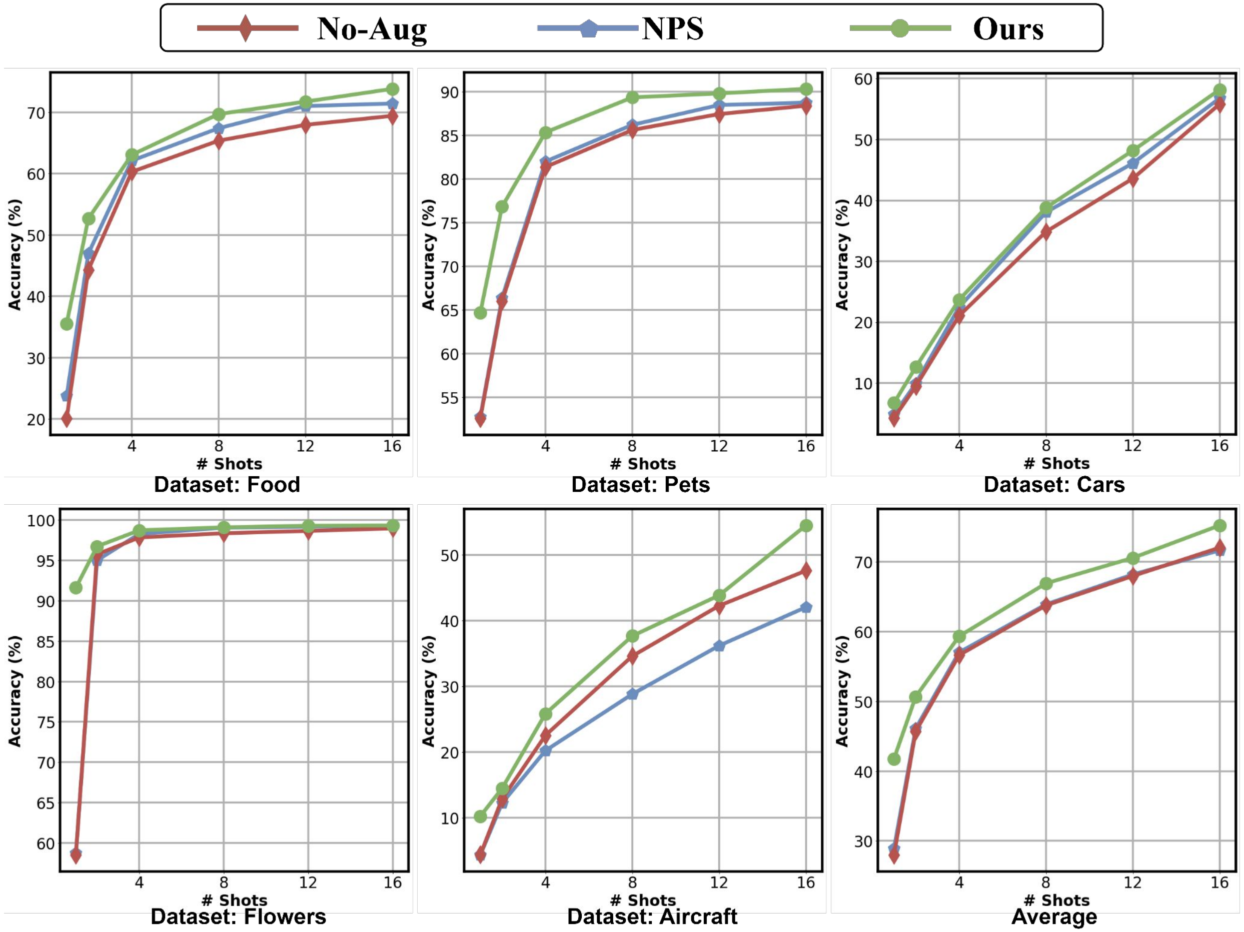
Hint-Aug: Evaluation Results

- A **2.22%** higher accuracy with **50%** less training data on Pet dataset



Hint-Aug: Evaluation Results

- A **2.22%** higher accuracy with **50%** less training data on Pet dataset
- **+0.04%~+32.91%** higher accuracy across different shots, tuning methods and datasets



JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Efficient and Intelligent Computing Lab

Georgia Tech

Hint-Aug: Drawing Hints From Foundation Vision Transformers Towards Boosted Few-Shot Parameter-Efficient Tuning

Zhongzhi Yu, Shang Wu, Yonggan Fu, Shunyao Zhang, and Yingyan (Celine) Lin

Georgia Institute of Technology

Paper Tag: WED-AM-273

The work was supported by the National Science Foundation (NSF) through the NSF CCF program (Award number: 2211815) and supported in part by CoCoSys, one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

Project Page:

Scan Me