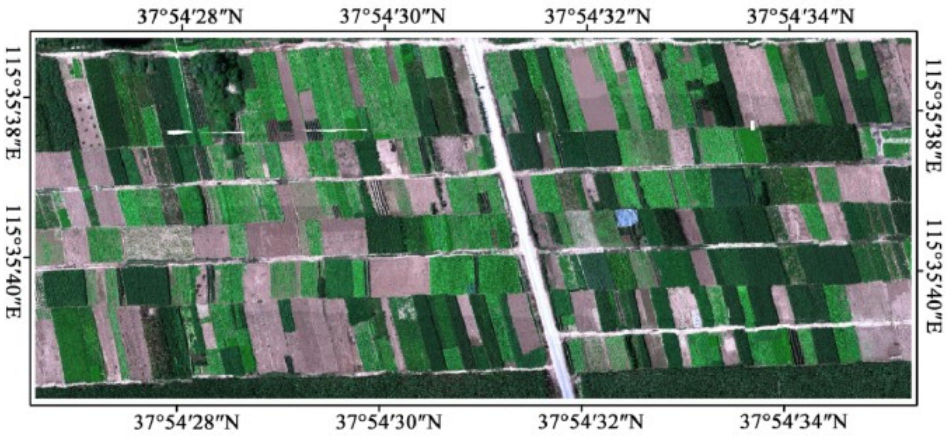
Improving Semantic Segmentation through Task Adaptation for UAV Hyperspectral Agricultural Imagery

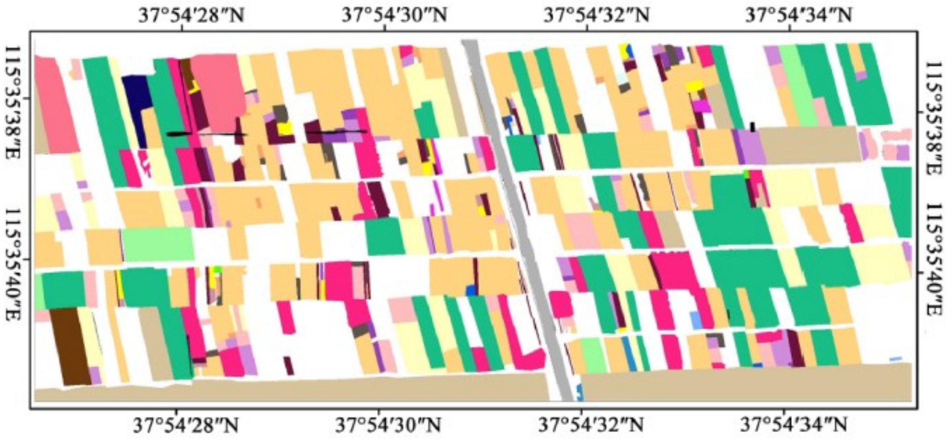
Mazharul Hossain, CS Ph.D. Student
Dr. Aaron L Robinson, Assoc Professor, EECE
Dr. Lan Wang, Professor, CS
Dr. Chrysanthe Preza, Professor and Chair, EECE

Date: **April 14, 2025**

Introduction

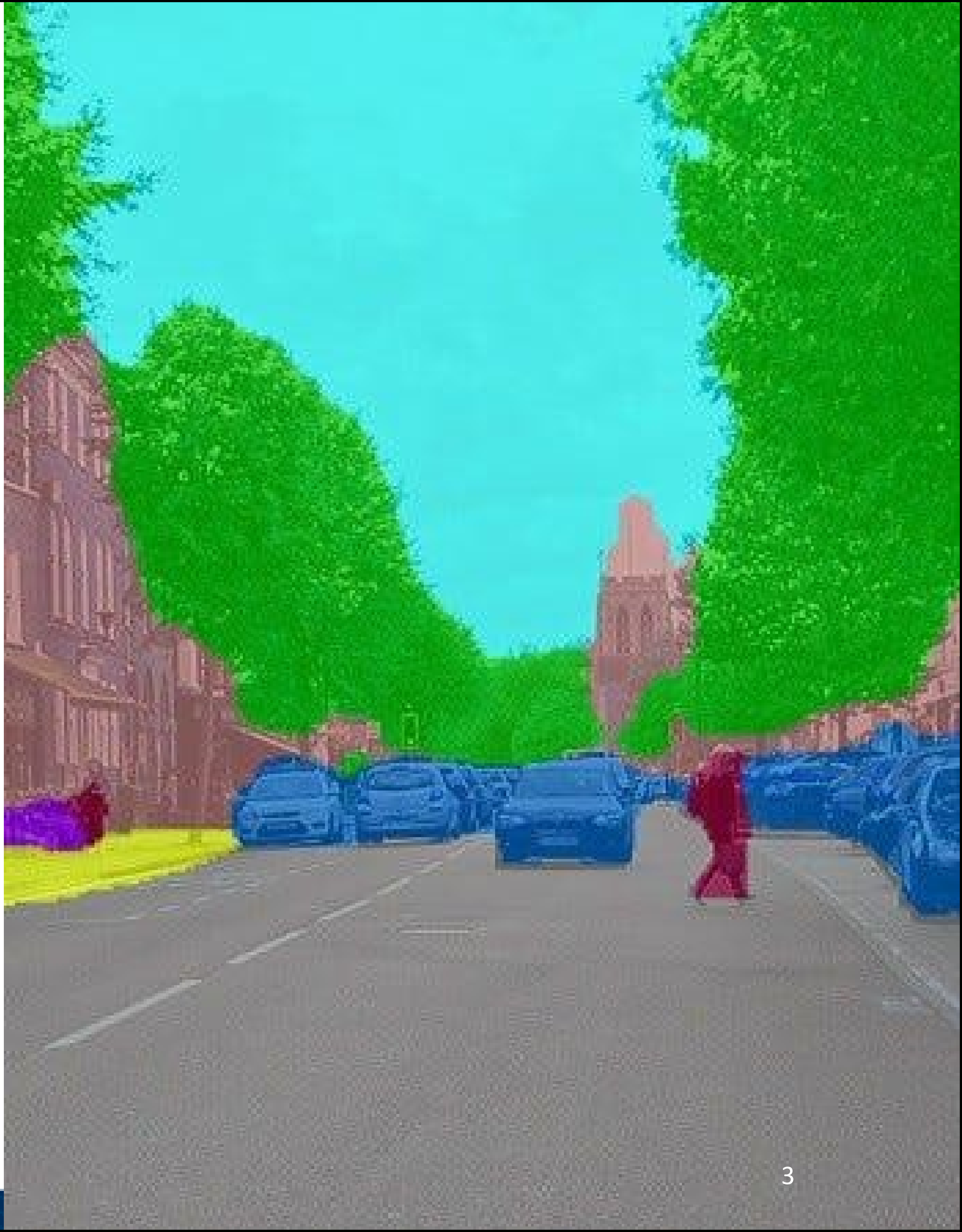
- **Why Crop Mapping Matters:**
  - Identifies crops, estimates yields, helps monitor plant health, and improves resource efficiency
- **Traditional methods** are slow and costly
  - What if AI could do it faster and more accurately?
- **Solution?**
  - Better accuracy: Deep learning
  - Faster: Remote sensing using hyperspectral imaging





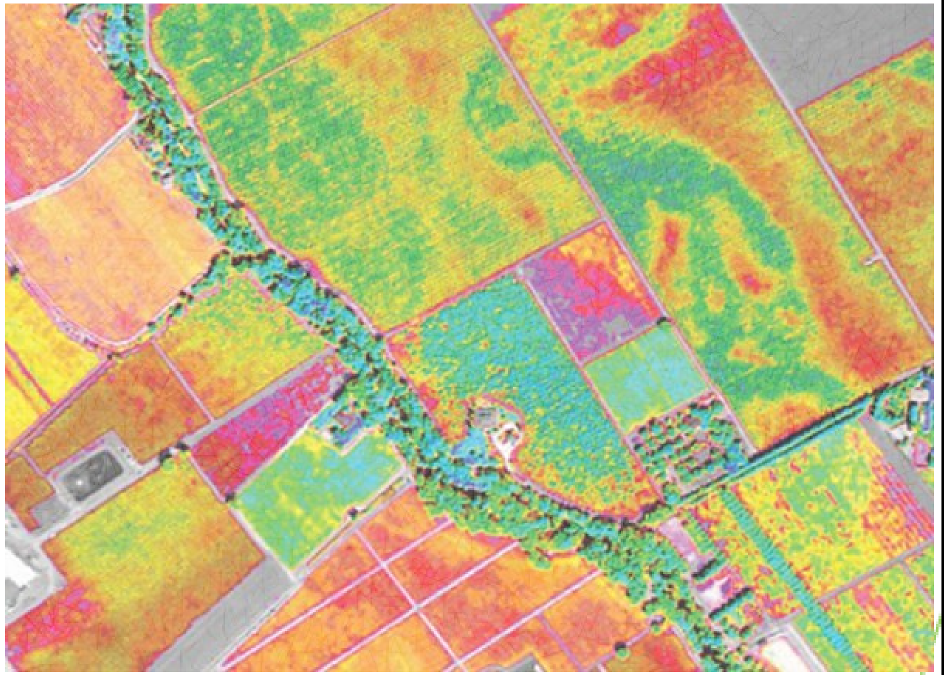
The Role of Semantic Segmentation

- **Definition:** Classifies each pixel in an image to differentiate objects using a deep learning (DL) algorithm
- **Applications:** Used in agriculture, medical imaging, autonomous vehicles, and industrial quality inspection.
- Existing models
  - Deep convolutional neural networks: U-Net, FastFCN, DeepLab
  - Transformer-based models: Segmenter, ViT (VisionTransformer).

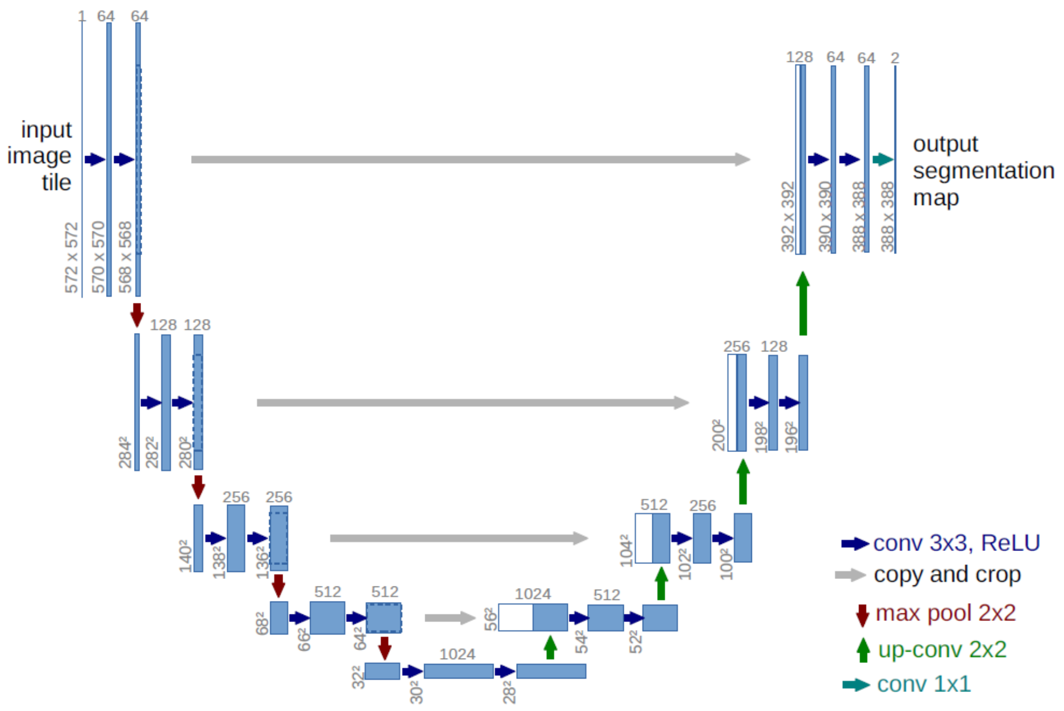


Hyperspectral Semantic Segmentation

- **Why Hyperspectral Over RGB?**
  - More spectral data = better differentiation of crops and weeds.
  - Near-Infrared (NIR): additional information a plant emits
- **Existing AI models struggle with hyperspectral images**
  - High-dimensional data
  - Limited annotated datasets
  - Deep classifiers overfit
  - Segmentation models struggles even more
- **Our Solution:** HSI-ResNetV2-ViT-Unet (HRViTUNet)
  - **Task Adaptation:** adapt pre-trained RGB segmentation models to hyperspectral data and agricultural domain
  - **Attention module**: reduce number of feature channels to improve efficiency



Related Work



- nnU-Net[1] (no new Unet) introduced an adaptive framework for vanilla U-Net
- TransUNet[2] replaced the bottleneck layer with ViTs, demonstrating promising results in medical imaging
- HSI-TransUNet[3] modified TransUNet with attention module

Our Proposed Solution

- **Task Adaptation:** Used pre-trained RGB models to improve hyperspectral segmentation and kept modification to a minimum

- Updated the input layer to handle different input modalities (HS imagery)

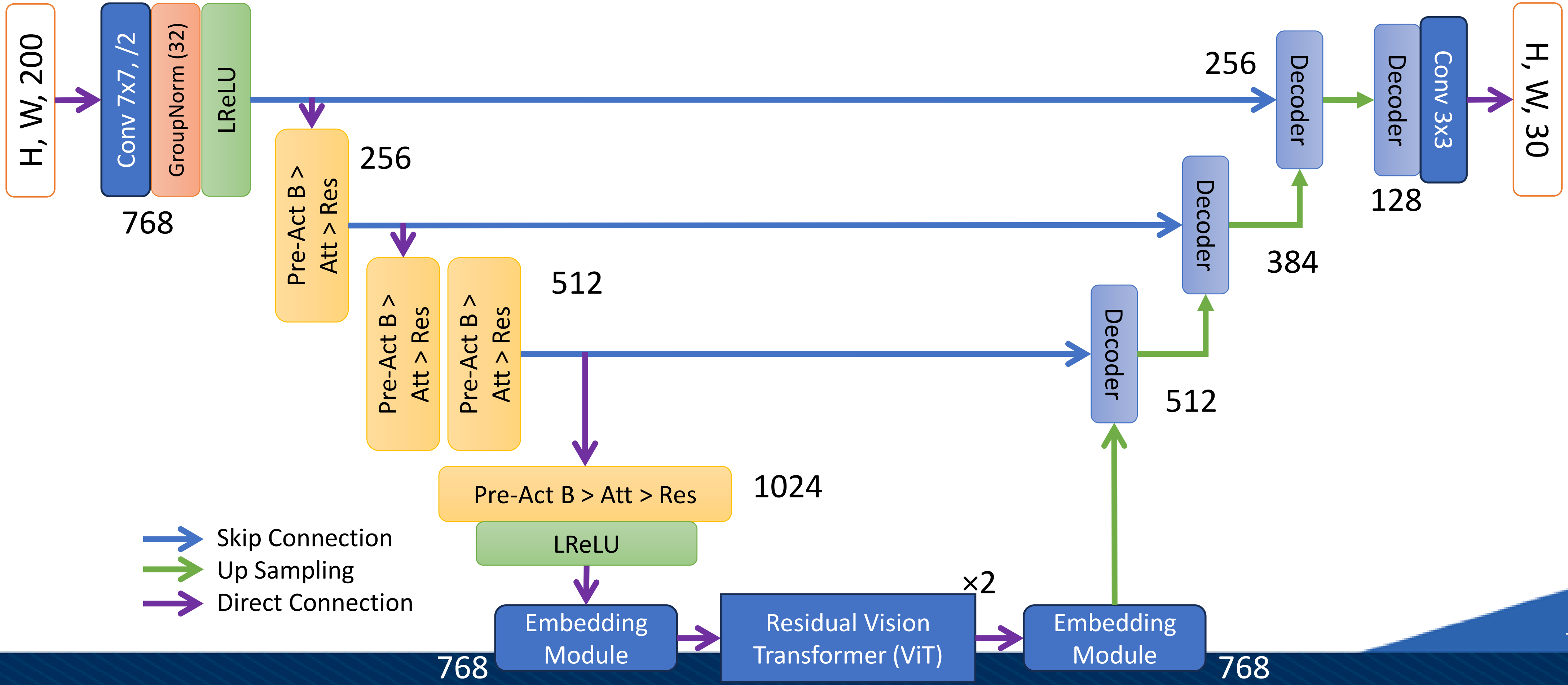
$$channel_{out} = 3 * 2^{\log_2 channel_{in}}$$
$$channel_{out} = 3 * 2^{\log_2 (200)}$$

- Updated the final input layer to handle increased number of classes

$$channel_{in} = 2^{\log_2 (channel_{out} * 3)}$$
$$channel_{in} = 2^{\log_2 (30 * 3)}$$

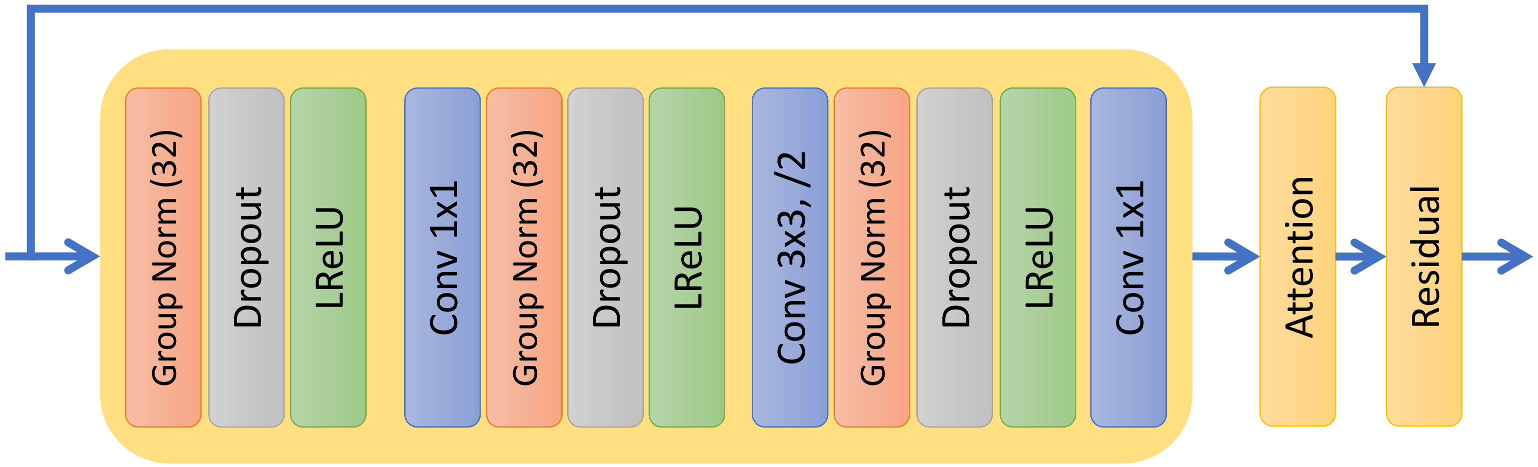
Our Proposed Solution (2)

- **ML Model Used:** Adapted a U-Net-like network with a spectral attention mechanism



Our Proposed Solution (3)

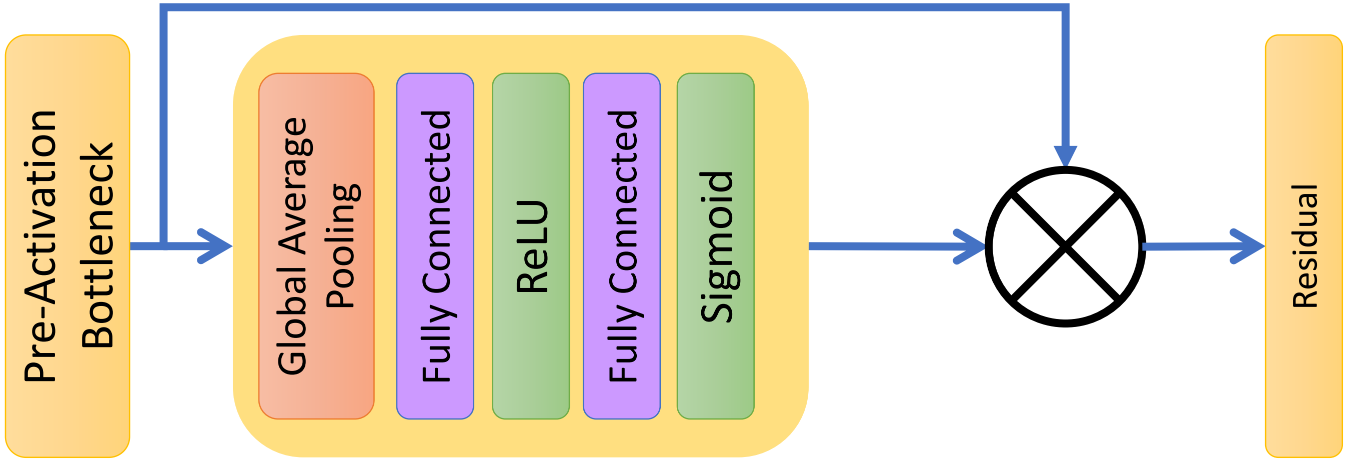
- **Key Enhancement 1:** Updated Pre-Activation Bottleneck block of ResNetV2
  - Better transmission of meaningful spectral features



Pre-Activation Bottleneck

Our Proposed Solution (4)

- **Key Enhancement 2:** New Adaptive Channel Attention Mechanism[4] to extract meaningful spectral features

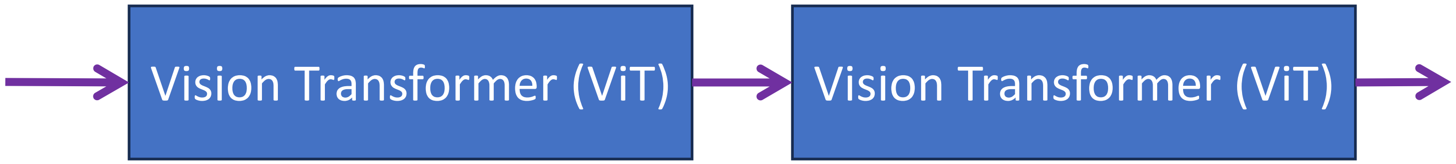


Adaptive Channel Attention Mechanism

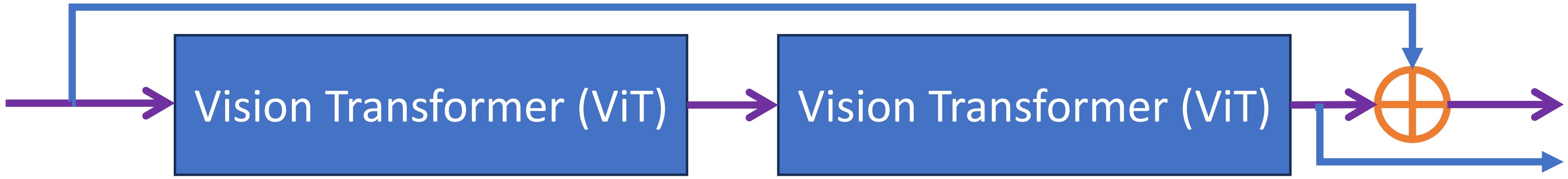
- We used an adaptive reduction ratio compared to a fixed 16

$$\text{adaptive reduction ratio} = 2 * log_2(C_{out})$$

Our Proposed Solution (5)

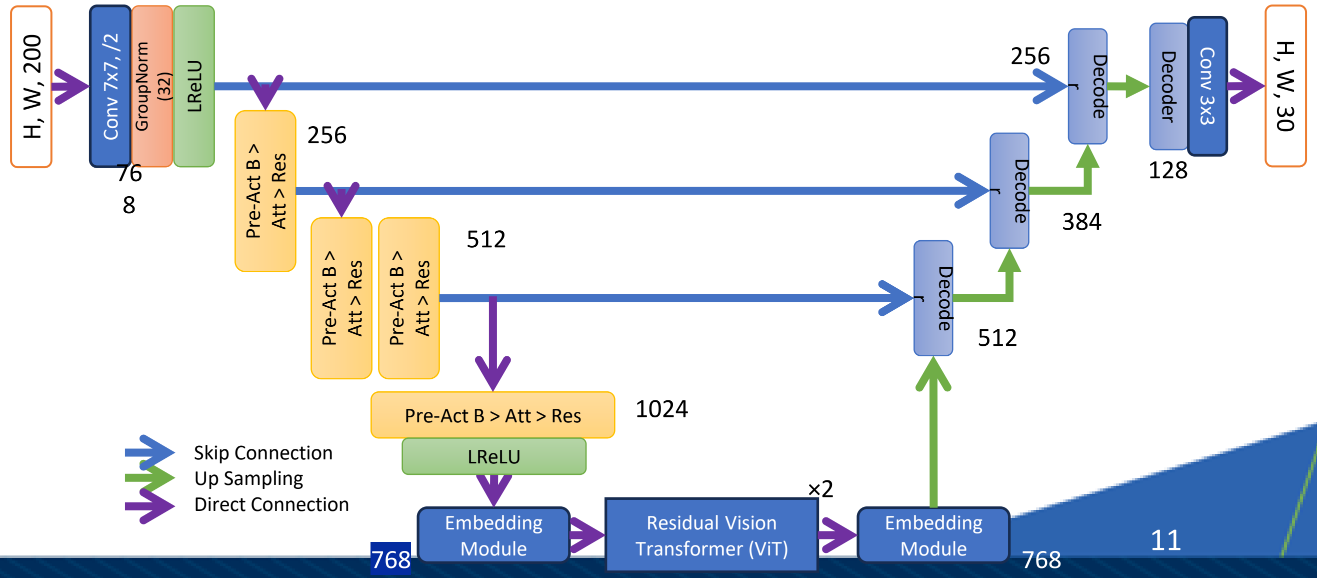


- **Key Enhancement 3:** alternating skip connections in Vision Transformer (ViT)
  - Avoids vanishing gradient
  - Improves backward gradient flow, efficiency, and performance



How It Works

- **Step 1:** Use ResNetV2 and ViT trained on large RGB image dataset
  - Pre-trained on ImageNet-21k dataset and fine-tuned on ImageNet dataset
- **Step 2:** Modify the input of the model to accept hyperspectral images
- **Step 3:** Modify the output of the model for new classes
- **Step 4:** Use channel attention to emphasize key spectral details
- **Step 5:** Fine-tune the whole model



Training Process Overview

- Loss Function
  - *Dice + Cross-Entropy* + Weighted *Cross-Entropy*
  - Combines pixel-level accuracy with overlap quality
- Optimizer & Learning Rate: *AdamW* (initial LR=0.000_1)
  - Reduced based on training down to 0.000_001

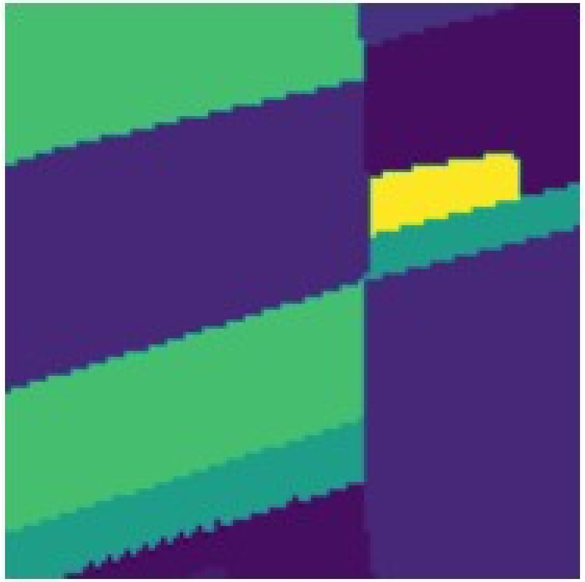
- Epochs: 500 (batch size: 32)
  - Early stopping (Dice)
- Data augmentation: PyTorch transforms (Flip, Rotate)
- Regularization: Dropout, Weight decay
- Hardware: 48GB NVIDIA RTX 6000 Ada Gen

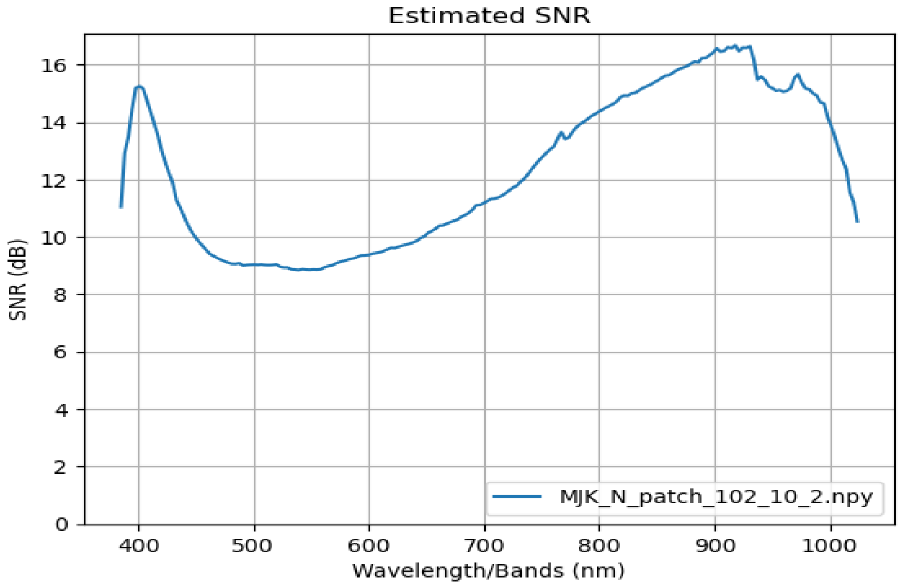
Input Image batch → Model → Loss → Optimizer → Updated Weights (20s/epoch)

Dataset

- HSI UAV crop dataset: collected from: Hebei Province, China
  - Collected using: Pika-L HS imager
  - Image size: 96*96*200 and Class: 30

Dataset	Training	Validation	Test
UAV-HIS-Crop Dataset	363	33	37





Qualitative Performance Evaluation

(a) RGB	(b) Ground Truth	(c) Our HRViTUNet (74.9 D)	(d) TransUNet (RGB) (77.9 D)	(e) TransUNet (HSI) (76.6 D)

The visualization shows a comparison of semantic segmentation between our model and TransUNets, where all models are trained and fine-tuned on the UAV-HIS-Crop Dataset.

Quantitative Performance Evaluation

Performance comparison of HRViTUNet using segmentation metrics

Model	Dice (mean, median)	Jaccard (mean, median)	Params (M)	Tflops (Tera)
HRViTUNet (Our)	0.749 ± 0.119, 0.749	0.737 ± 0.123, 0.732	**57.70**	2.19
TransUNet (RGB)	**0.779 ± 0.097, 0.794**	**0.771 ± 0.099, 0.785**	105.16	111.58
TransUNet (HSI)	0.766 ± 0.097, 0.760	0.757 ± 0.099, 0.753	105.78	111.65
HSI-TransUNet	0.643 ± 0.133, 0.648	0.631 ± 0.137, 0.633	99.33	**1.18**

Performance comparison of HRViTUNet using classification metrics

Model	Precision	Recall	F1-score	Accuracy
HRViTUNet (Our)	**0.334 ± 0.166**	0.552 ± 0.201	**0.277 ± 0.159**	**0.766 ± 0.121**
TransUNet (RGB)	0.315 ± 0.171	0.409 ± 0.238	0.226 ± 0.168	0.606 ± 0.248
TransUNet (HSI)	0.315 ± 0.154	0.470 ± 0.231	0.242 ± 0.155	0.668 ± 0.232
HSI-TransUNet	0.258 ± 0.141	**0.554 ± 0.216**	0.205 ± 0.140	0.746 ± 0.185

Summary

- We applied **Task Adaptation** to a pre-trained RGB model to improve crop mapping efficiency and accuracy with hyperspectral data
  - **Attention** module
  - Custom **loss** function
- **Comparison with other models**
  - Comparable in segmentation performance
  - Better in classification performance
  - Much lower computational cost
- With advancements like these, AI can revolutionize how we manage agriculture, making farming smarter and more efficient

Real-World Impact and Next Steps

- **Farmers & Remote Sensing Experts** for more effective decision-making
  - More accurate crop mapping
  - Crop monitoring
  - Enhanced analysis of land use
- **AI Researchers:** New applications for hyperspectral image processing
- **Dual-Use Potential:** Can aid in segmenting and classifying anomalies into objects of interest, enhancing surveillance and situational awareness

- **Next Steps:**
  - Test on other hyperspectral datasets.
  - Use Unmixing to move task adaptation to the next step and reduce training even further
  - Optimizing model efficiency for real-time applications

References

1. Isensee, Fabian, et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature Methods 18.2 (2021): 203-211.
2. Chen, Jieneng, et al. "Transunet: Transformers make strong encoders for medical image segmentation." arXiv preprint arXiv:2102.04306 (2021).
3. Niu, Bowen, et al. "HSI-TransUNet: A transformer based semantic segmentation model for crop mapping from UAV hyperspectral imagery." Computers and Electronics in Agriculture 201 (2022): 107297.
4. Hu, Jie, et al. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

Thank you