# Improve U-Net Image Segmentation efficiency with LoRA Fine tuning

*Yunxiang Wang, Jiawei Guo*

## Motivation

- **Semantic segmentation** is used in various tasks.
- High-performing models requires **large datasets** and **high computational resources**.
- Data collection annotation is **labor-intensive** and **complex**.
- Limited access to high-performance hardware **restricted** training efficiency
- LoRA(Low Rank Adaptation) shows promise in **reducing computational costs**
- Application of it in semantic segmentation are **underexplored**

## Dataset

- **GTA5 Dataset** (from "*Playing for Data: Truth from Computer Games*")
- Modeled from Los Angeles urban environment
- Contains 24966 **high-resolution** images (1914x1052 pixels)
- **Pixel-level** annotations for 35 semantic categories

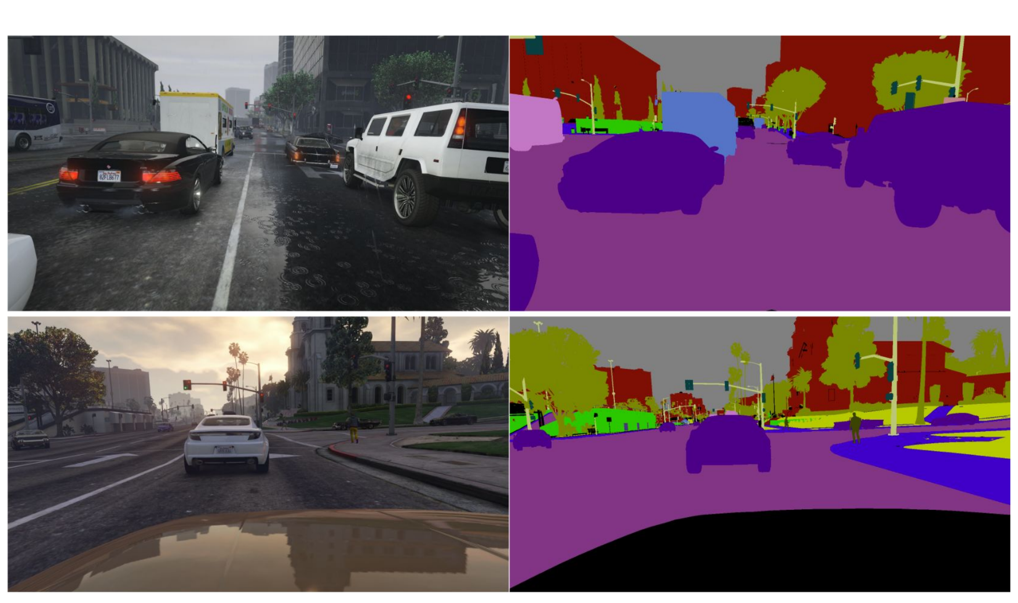

*Figure 1: Example urban scenario images from GTA5 Dataset. Left: original image. Right: data with annotations including car, trucks, tree, building, road, human, etc…*

- Convert from PNG (~2MB) to JPG (700KB)
- Random crop to 512-512 pixels during train
- Convert RGB mask to index mask
- Train set: 6,500
- Test set: 1,000

## References


- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-rank adaptation of large models. *arXiv*. https://doi.org/10.48550/arXiv.2106.09685
- Richter, S. R., Vineet, V., Roth, S., & Koltun, V. (2016). Playing for data: Ground truth from computer games. *European Conference on Computer Vision (ECCV)*, 102-118. https://doi.org/10.1007/978-3-319-46475-6_7
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28


## Architecture

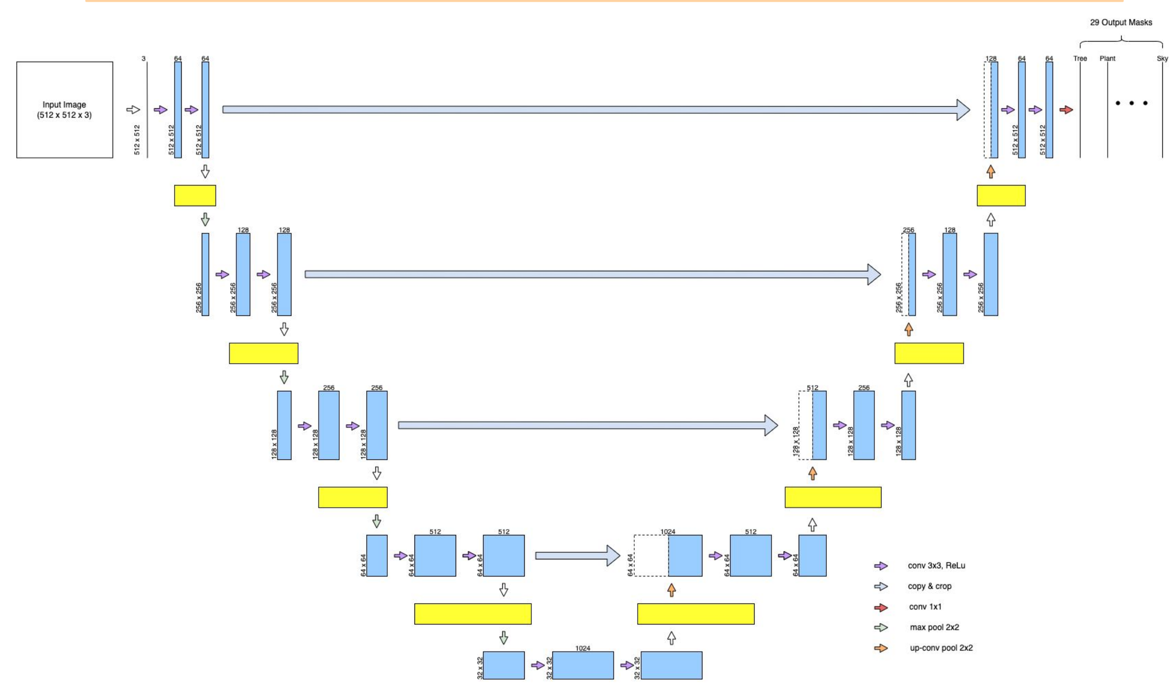

*Figure 2: Modified U-Net Structure with LoRA applied after each convolution layers. .*

- Leveraged the traditional U-Net structure
- Multiclass Cross Entropy loss
- Apply LoRA on all convolution layers
- Freeze all parameter of U-Net during training, only update the appended LoRA's weight

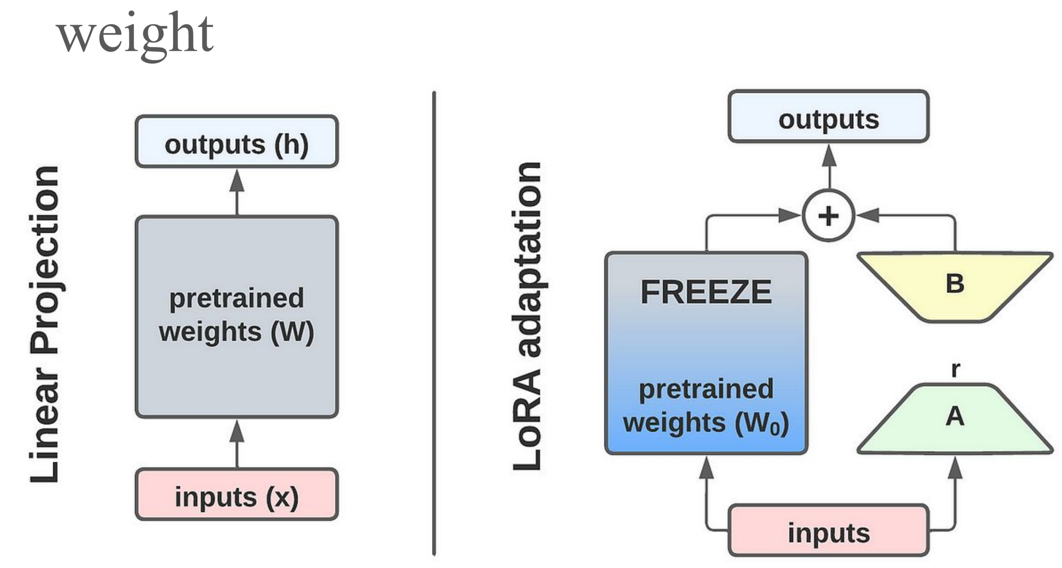


*Figure 3: Workflow of Low Rank Adaptation (LoRA). .*

## Experiment

- First, we implement U-Net segmentation and LoRA from scratch.
- Then we tested our model on binary-class segmentation dataset
- Achieved **97%** accuracy
- But overfitted to specific features and poor generalization.
- Therefore, we decided to test on multiple classes GTA5 dataset.
- We planed to first train our baseline U-Net model to have a stable weight as initialization.
- Then we trained 2 models separately after the original checkpoint.
- We recorded the training loss, training time per batch, and check validation loss after each epoch.
- At the same time, we monitored computational resources.

## Future Work

- What are the performance difference between LoRA and other techniques such as transfer-learning
- How does LoRA perform in fine tuning of other model?

## Result

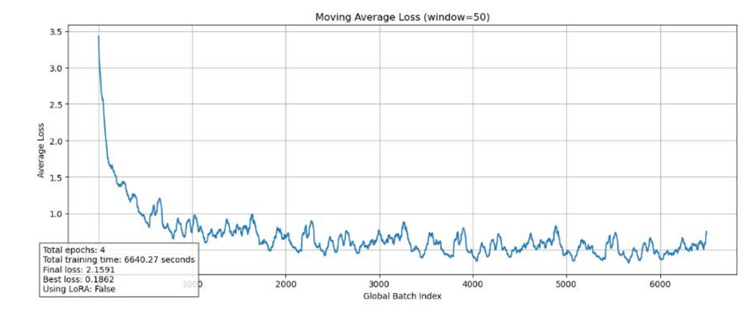

*Figure 4: Moving Average Loss of baseline model without LoRA, with best loss 0.1862*

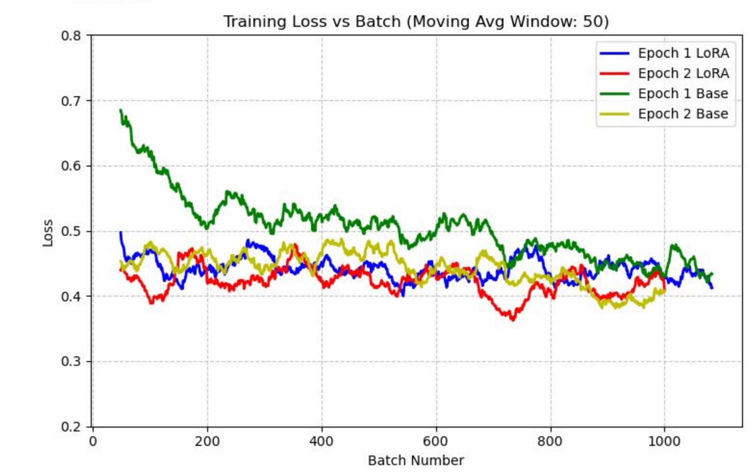


*Figure 5: Trend of training loss over batch size. LoRA with overall low loss.*


*Figure 5: Trend of training loss over time. LoRA helps reduces training time.*

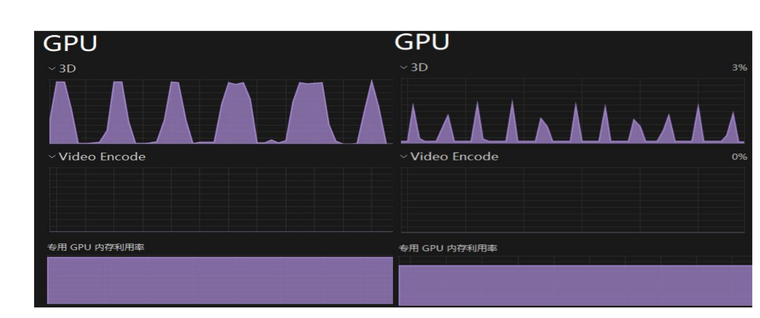


*Figure 6: Train on RTX 4070. Left image training regular U-Net. Right image training with LoRa*

| Model | Accuracy | Mean Loss | Training time | Trainable Parameters |
|---|---|---|---|---|
| Baseline | 72.36 | 0.6722 | 75.47 | 35.5 |
| Model1 | 81.12 | 0.4938 | 87.52 | 0.42 |
| Model2 | 82.57 | 0.4438 | 80.33 | 0.42 |

*Table 1: Model comparison result. Training time in minutes. Trainable Parameters in million.*

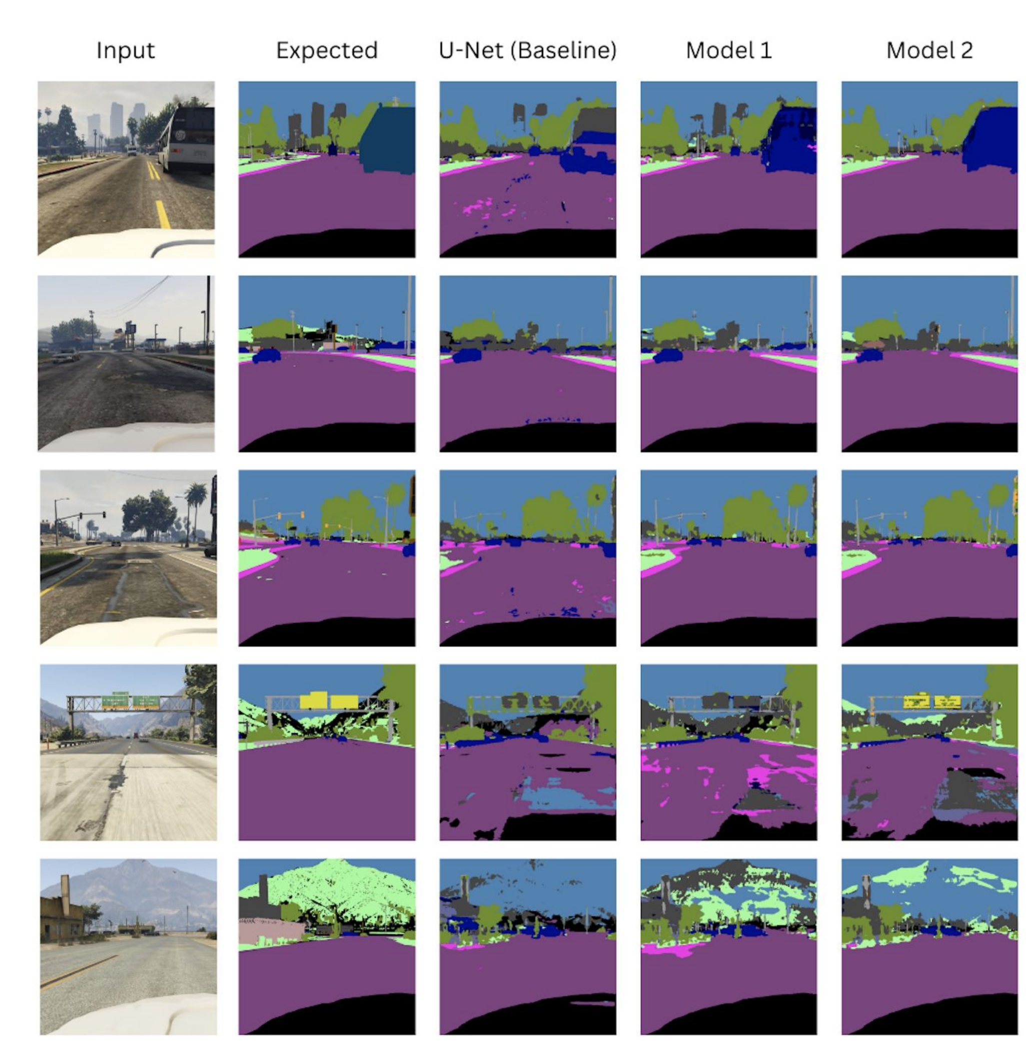


*Table 2: Segmentations on validation data compare with Baseline U-Net, ground truth, and original image. Model 1: with LoRA. Model 2 with additional data augmentation.*

## Conclusion

- LoRA model sees improvements in average training loss, accuracy, GPU occupancy and memory usage
- Observed significant acceleration in training time (~20%)
- We demonstrated LoRA's speed advantage in training, but cannot attribute it solely to LoRA since training times fluctuate between epochs, particularly after the first epoch
- LoRA model perform better visually in several small categories, such as poles, frames, mountains…