Mid-term Report:

1. Current Progress

- **Theoretical & Env Prep**: Analyzed the paper's core improvements (ResMLP, URIM, DF loss), built a matching hardware (3.8GHz CPU, 8-core 32G+1 V100 GPU) and software (PyTorch) environment, and preprocessed the Ma et al. public dataset (20 volume images → 2000 augmented images, 8:2 split) .
- **Preliminary Results**: Completed & trained the original Swin-Unet, achieving Dice≈0.779, Precision≈0.775, Recall≈0.785, IOU≈0.638 (consistent with the paper). Integrated ResMLP into the model, which runs normally .

2. Key Difficulties

- **Module Implementation**: URIM's LC-Conv details (e.g., kernel size) are unclear; DF loss weight tuning needs experiments .
- **Dataset Issues**: Low image resolution, insufficient diversity, and potential annotation differences affect model training/evaluation .
- **Efficiency**: Each training round takes ~1.5 hours; multi-experiment & metric calculation are time-consuming .

3. Proposal Adjustments

- **Phased Module Integration**: First ResMLP, then DF loss, finally URIM .
- **Dataset & Env Opt**: Add Gaussian noise/flip for augmentation; use checkpoints & multi-GPU training .
- **Evaluation Upgrade**: Add qualitative visualization (segmentation contours) and scale-region metric distribution analysis .

4. Key Q&A

- **SOTA**: U-Net variants (e.g., M2UNet) lack multi-scale ability; pure Transformers (original Swin-Unet) have weak local extraction; some models are parameter-heavy .
- **Re-implementation Reason**: The paper's model solves prior flaws (ResMLP reduces feature loss, URIM fixes uncertain regions, DF optimizes small targets) .
- **Evaluation**: Quantify via Dice/IOU/Hd; qualify via visualization. Compare with U-Net/Unet++/TransUnet/original Swin-Unet on Ma et al.'s dataset .

5. Team (Individual Statement)

This project is completed individually.