# SuperRivolution: Fine-Scale Rivers from Coarse Temporal Satellite Imagery

Supplementary Material

## A. Additional Results

### A.1. Segmentation performance.

**Segmentation outputs compared to baselines.** Fig. A1 shows visualization of segmentation outputs from different methods. The HR model (RiverScope) outputs segmentations with precise boundaries and are very close to the ground truth labels. In contrast, the sentinel baseline outputs ragged segmentations with less defined boundaries. Input upsampling is generally more consistent with the ground truth than output upsampling and super-resolution. Output upsampling tends to miss fine-grained details, which is likely due to the upsampling only being done after the model has already made its predictions—the model is operating on LR images, which naturally loses details. Super-resolution, on the other hand, can hallucinate details (last row of Fig. A1) causing it to make mistakes.

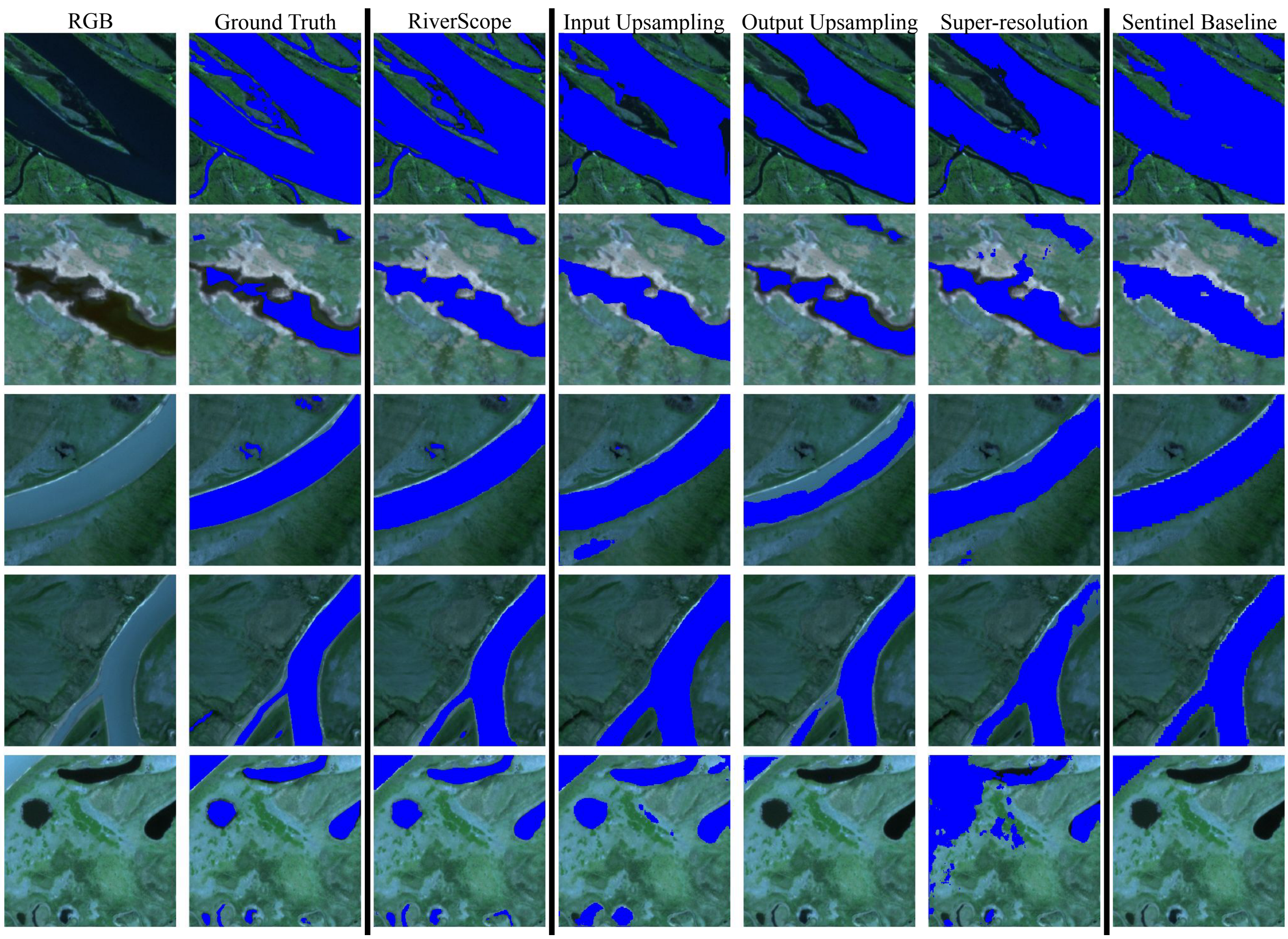


Figure A1. **Visualization of segmentation outputs from different methods**.

**Segmentation comparison of single and temporal inputs.** Fig. A2 shows the qualitative comparison of different methods that use single or temporal inputs. In general, more details can be distinguished when using temporal inputs, resulting in better segmentation quality. Using temporal inputs helps the models be less susceptible to atmospheric noise like clouds.

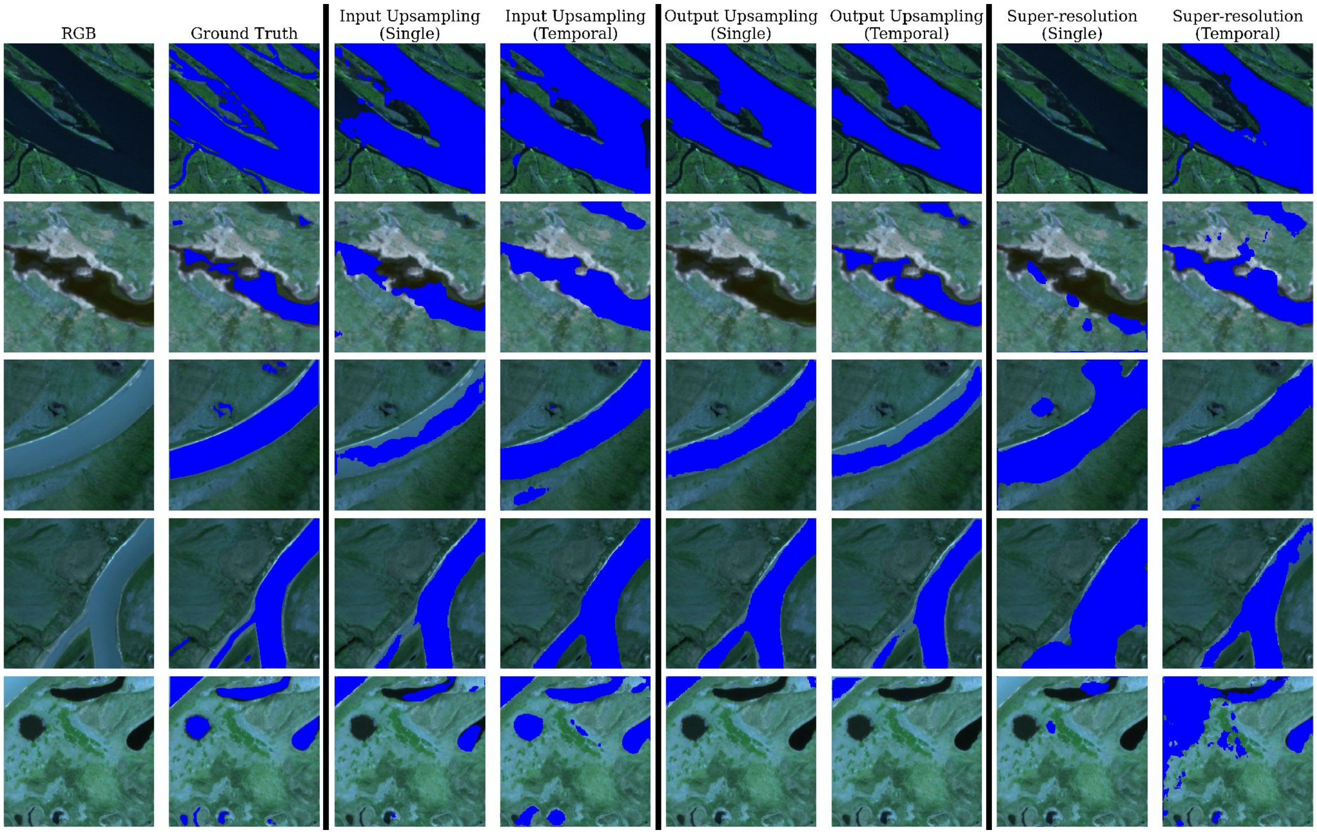


Figure A2. **Visualization of segmentation outputs when using single versus temporal inputs**.

**Results across pretraining methods.** Fig. A3, Fig. A4, and Fig. A5 show segmentation performance across different pretraining methods, segmentation models, and backbones. Similar to trends observed in Fig. 3, using temporal inputs with SuperRivolution result in better F1 score than the low resolution Sentinel baseline (horizontal markers).

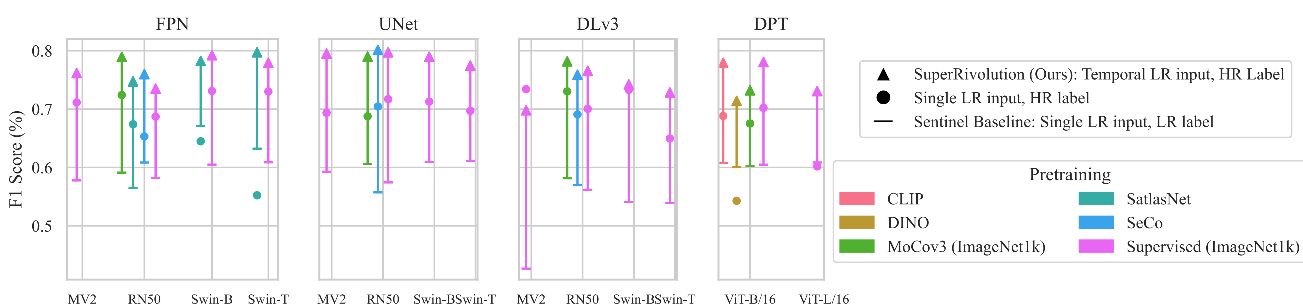


Figure A3. **Input upsampling segmentation metrics across different pretraining methods**.

## A.2. Cloud cover impact on performance

**Results across pretraining methods.** Fig. A6, Fig. A7, and Fig. A8 show the impact of performance when using a temporal input versus a single image input. The three plots show performance across different backbones, segmentation models, and pretraining methods. Similar to results in Fig. 4, we generally see high performance gain on cloudy images.

**Cloud cover impact on super-resolution.** Fig. A9 shows that using temporal input for an otherwise noisy or cloudy single image significantly improves the performance. In particular, when applying the super-resolution model ESRGAN [61] the

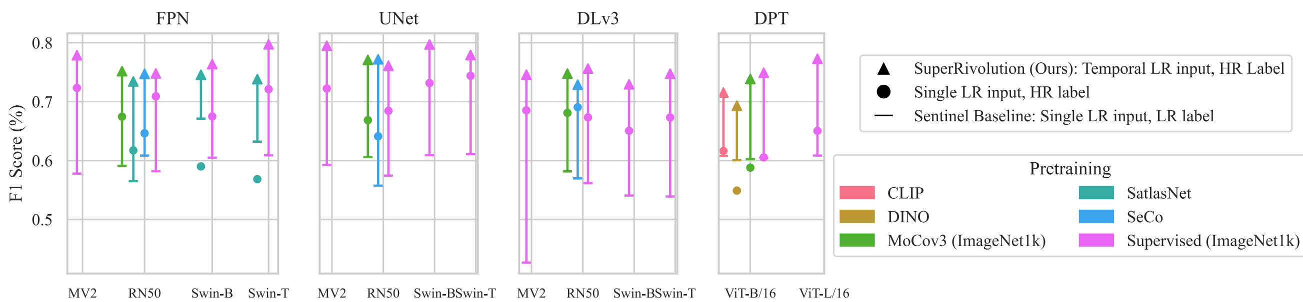


Figure A4. **Output upsampling segmentation metrics across different pretraining methods**.

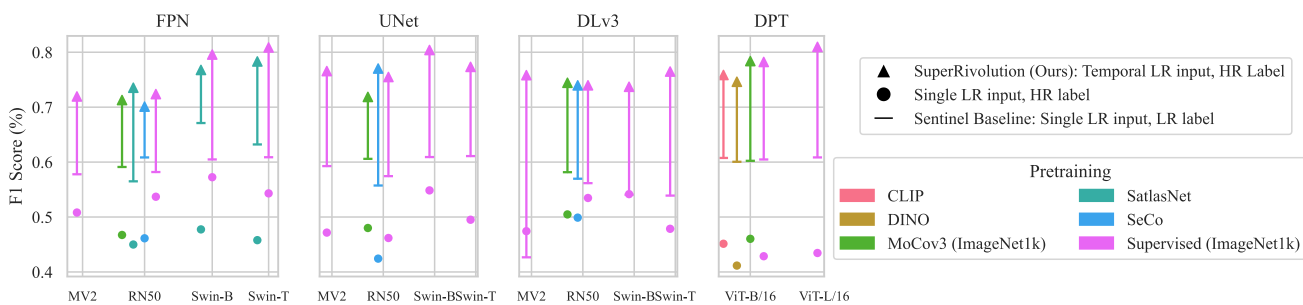


Figure A5. **Super-resolution segmentation metrics across different pretraining methods**.

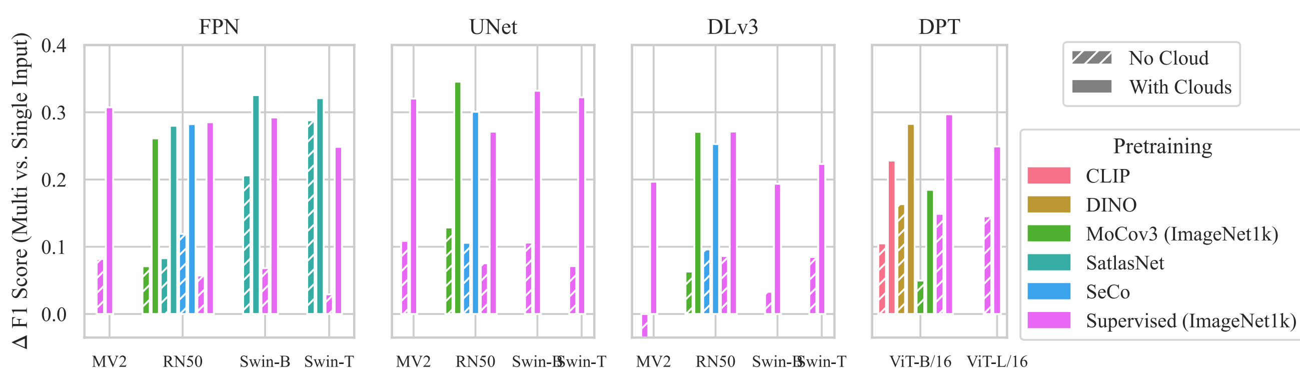


Figure A6. **Impact of cloud cover on temporal input performance gain for input upsampling**.

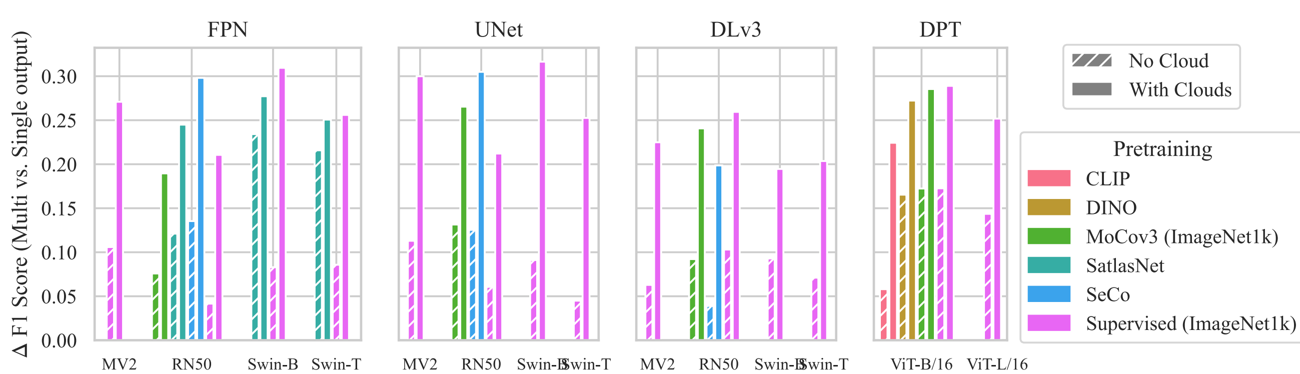


Figure A7. **Impact of cloud cover on temporal input performance gain for output upsampling**.

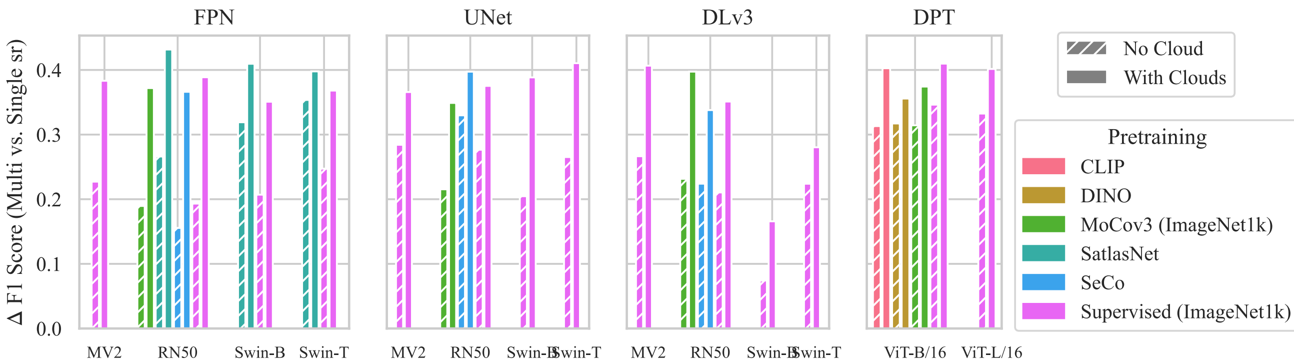


Figure A8. **Impact of cloud cover on temporal input performance gain for super-resolution**.

performance significantly improves for temporal inputs over single image inputs. By ensembling the model output across multiple inputs, the effect of the noise is minimized.

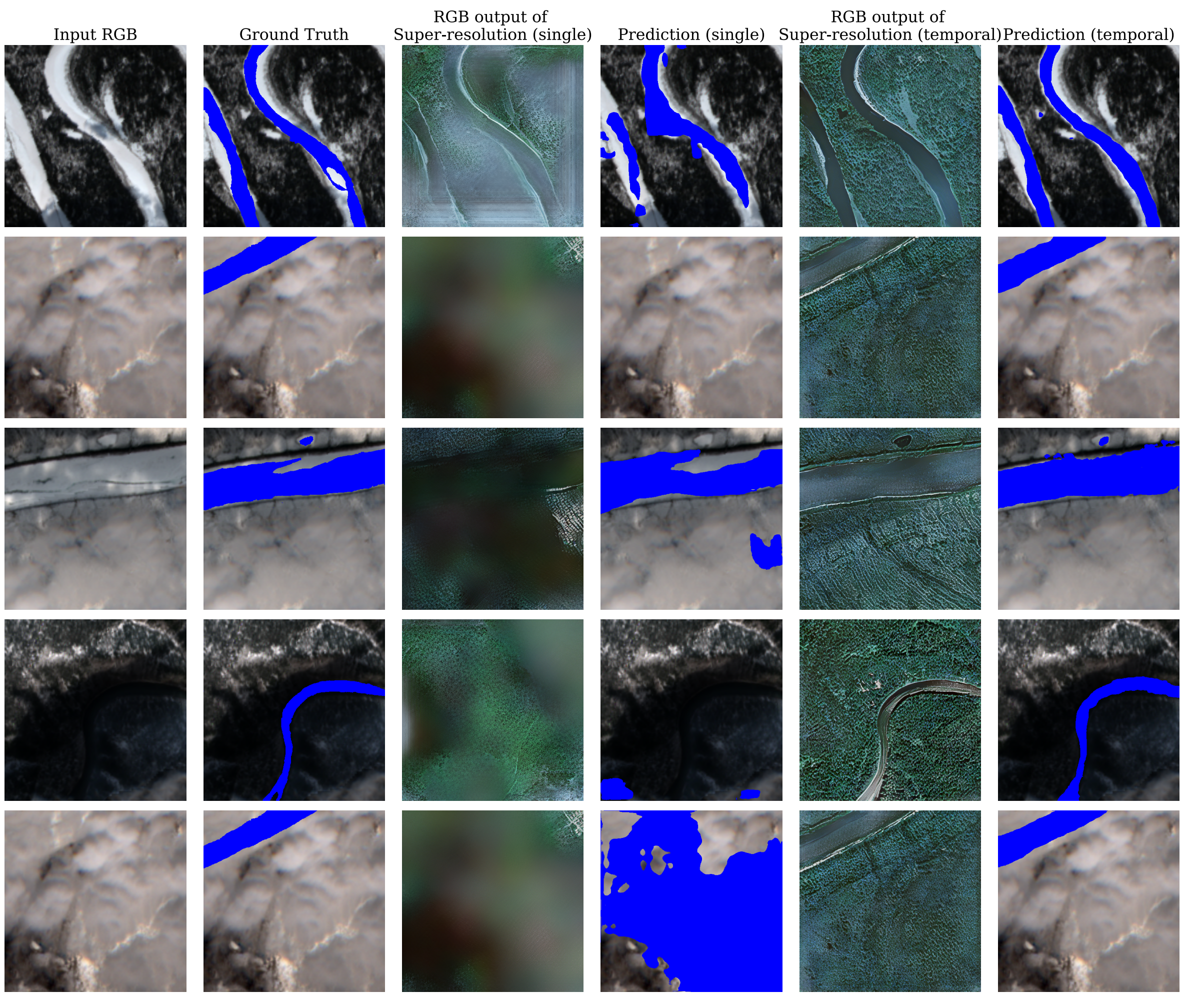


Figure A9. **Visualization of the impact of noisy and cloudy images when applying the super-resolution model**.

## A.3. River width estimation errors.

**Distribution of width estimates.** Fig. A10 shows the distribution of river width estimates of the different methods. Models that take multiple images as input are generally less biased with the prediction clustering closely to the $y = x$ line.

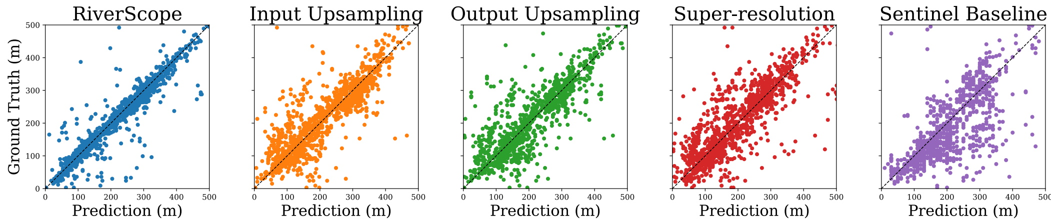


Figure A10. **Distribution of predicted widths from different methods**. Temporal inputs are used for input upsampling, output upsampling, and super-resolution. Single images are used for RiverScope and Sentinel Baseline. Sentinel baseline predictions appear to be more skewed than the SuperRivolution methods.

**Results across pretraining methods.** Fig. A11, Fig. A12, and Fig. A13 show the width estimation performance across different backbones, segmentation models, and pretraining methods. Consistent with the results in Fig. 5, the performance improves when temporal inputs are used as input.

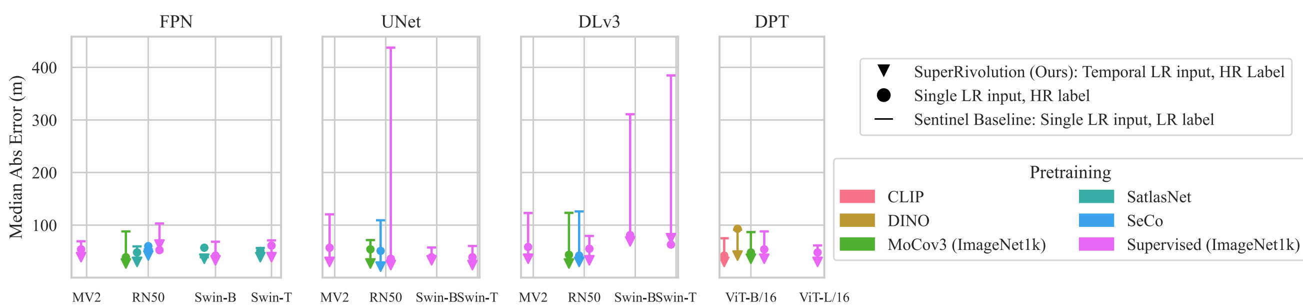


Figure A11. **Input upsampling river width estimation errors across different pretraining methods**.

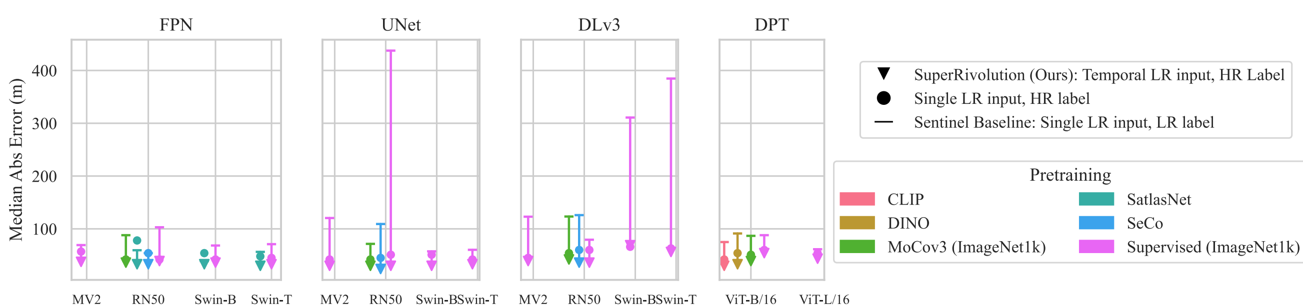


Figure A12. **Output upsampling river width estimation errors across different pretraining methods**.

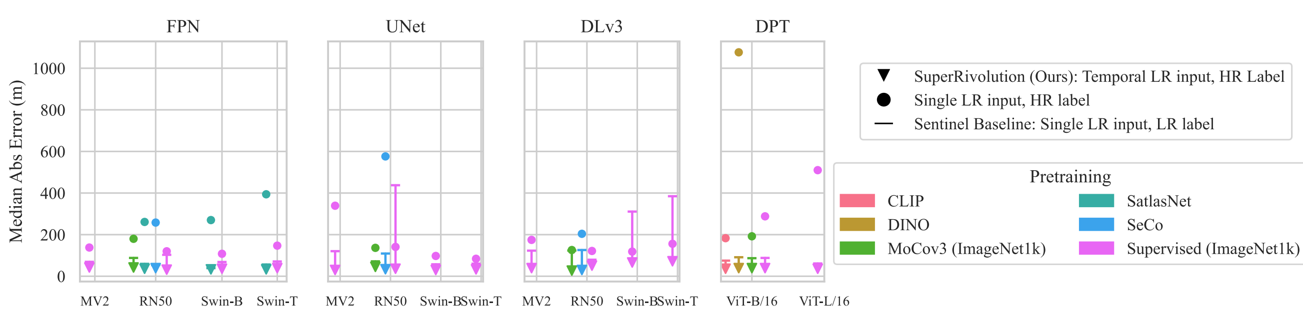


Figure A13. **Super-resolution river width estimation errors across different pretraining methods**.