H2O-Net: Self-Supervised Flood Segmentation via Adversarial Domain Adaptation and Label Refinement Supplementary Material

1. Additional Qualitative Results

Here we provide additional qualitative result to illustrate performance in each stage of H2O-Net. Figure 3 illustrates features SWIR-Synth learns in low resolution data, showing the input image, ground truth $SWIR_2$, synthesized $SWIR_2$, thresholded mask obtained from ground truth $SWIR_2$, and predicted thresholded mask obtained from synthesized $SWIR_2$. Figure 4 demonstrate SWIR-Synth qualitative performance on high resolution data. Figure 5 show refiner performance on low resolution data. Figure 1 shows additional results on PlanetScope [1], and figure 2 shows qualitative results on DroneDeploy [2].

2. Implementation Details

2.1. SWIR-Synth Network Architecture

SWIR-SynthNet is structured as an image-to-image translation network. The generator is constructed of 3 encoding blocks, and 3 decoding blocks with skip connections. Each encoding block is comprised of a spectrally normalized convolution layer with a 3×3 kernel and 1 padding, batch normalization layer, leaky rectified linear unit (ReLU) with slope of 0.2 and a dropout layer. Each block output is then down-sampled by a factor of 2 using a strided convolution layer. Decoding blocks first upsample input using a spectrally normalized transposed convolution layer with a 4×4 kernel size and a stride of 2, followed by a spectrally normalized convolution layer with a 3×3 kernel and 1 padding, batch normalization layer, leaky ReLU with slope of 0.2 and a dropout layer. Decoding blocks take a concatenation of the previous decoding block and output of the corresponding encoder block. The discriminator has 5 blocks each with a spectrally normalized convolution layer with 4×4 kernel size, and stride of 2, followed by a batch normalization, and a ReLU activation function. A self-attention module similar to [4] was added after the second and forth blocks. The last layer of the last block outputs the features tensor for the feature loss, which is also used as an input to a Sigmoid activation layer for the adversarial loss.

2.2. Segmentation Architecture

The segmentation network uses an encoder-decoder architecture similar to U-Net [3], with 5 encoding blocks and 5 decoding blocks with skip connections. Encoding blocks comprise of two spectrally normalized convolution layers, batch normalization layer, and LeakyReLU layer. Each block is followed by a strided convolution layer acting as a pooling layer. The decoder also has 5 blocks, each comprised of transposed convolution layer for upsampling, two spectrally normalized convolution layers, batch normalization layer, and a Leaky ReLU layer. Skip connections are used for corresponding encoding and decoding blocks. This architecture was also used for the U-Net baseline method to ensure consistency and fairness in evaluation.

2.3. Refiner Architecture

has one encoder block and one decoder block. The encoder block consists of 2 convolution layers each followed by a normalization layer, batch normalization layer after first convolution and instance normalization layer after second convolution, and a Leaky ReLU layer. The decoder block first upsamples features using bilinear interpolation, followed by 2 convolution layers, normalization layers, and Leaky ReLU layers. For high confidence points sampling, we use ϕ_H of 0.5, and ϕ_L of -0.2.

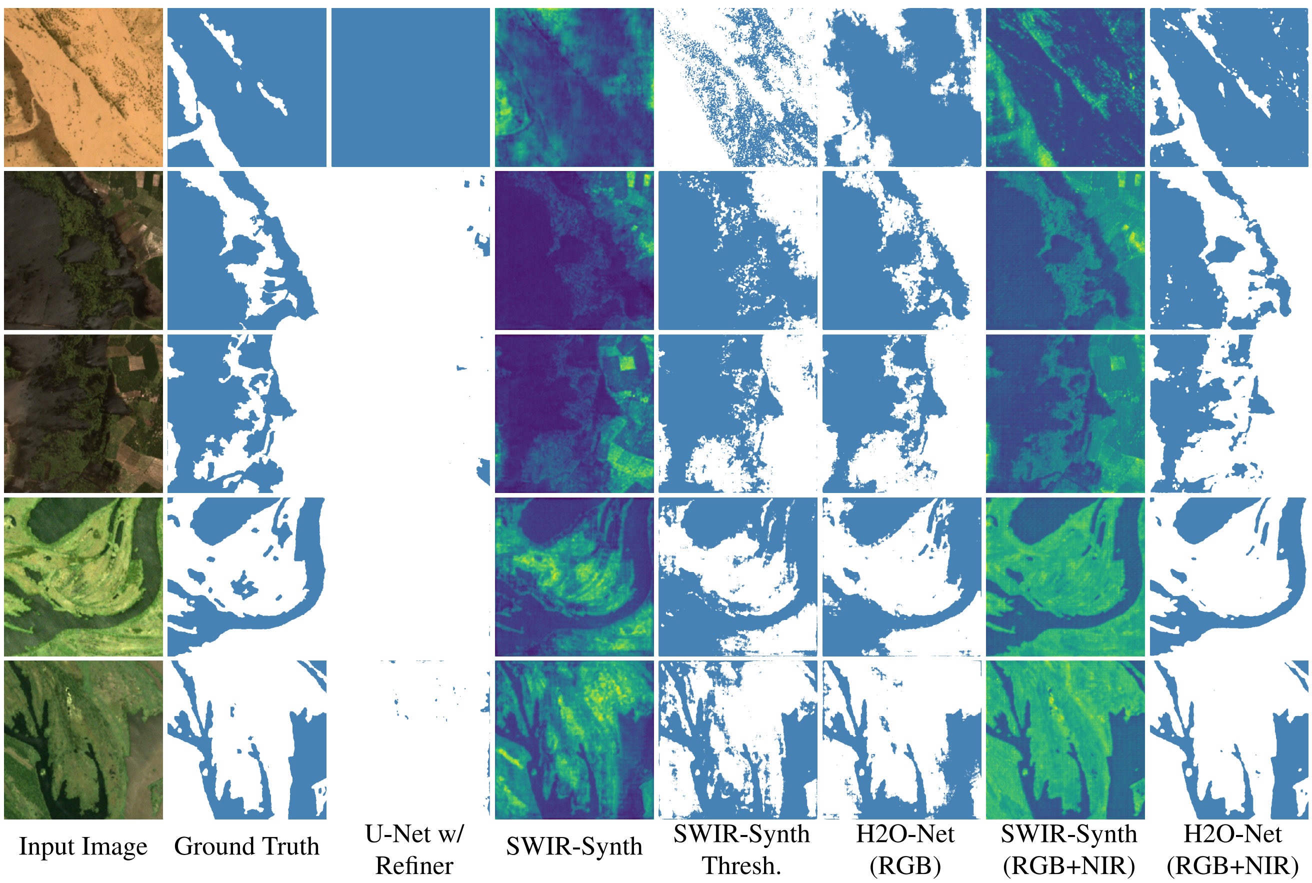

Figure 1: Qualitative comparison with SOTA methods on PlanetScope [1]. H2O-Net results shows that synthesizing SWIR data allows robust segmentation performance in high resolution imagery. It can be seen that adding NIR signals to training and inference improves results in both for segmentation and synthesized SWIR. Blue and white predictions correspond to water and non-water pixels. Best viewed in color and zoomed.

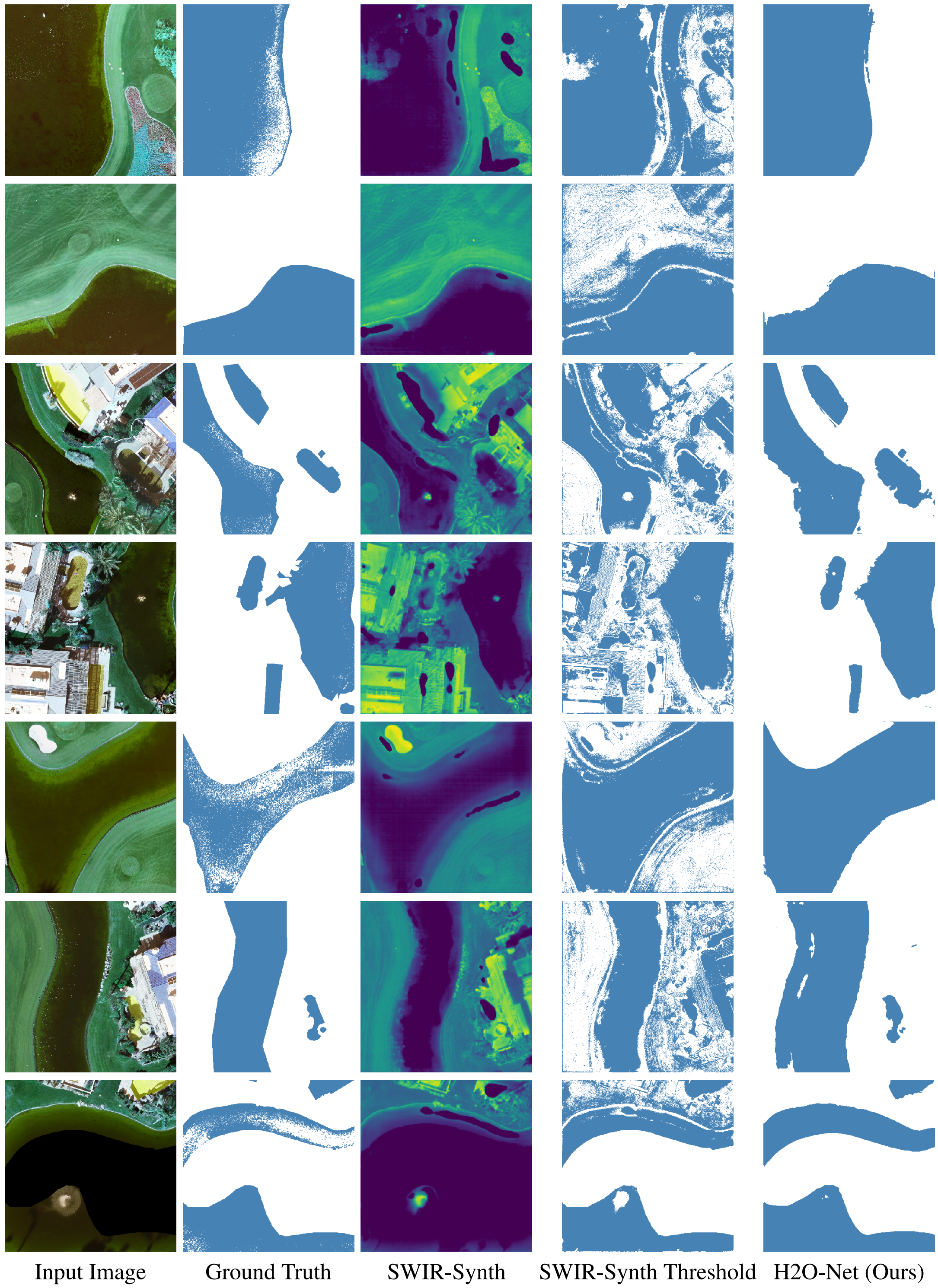

Figure 2: Qualitative comparison with SOTA methods on Drone Deploy. Our method (H2O-Net) shows that synthesizing SWIR data allows robust segmentation performance in high resolution imagery. Best viewed in color and zoomed.

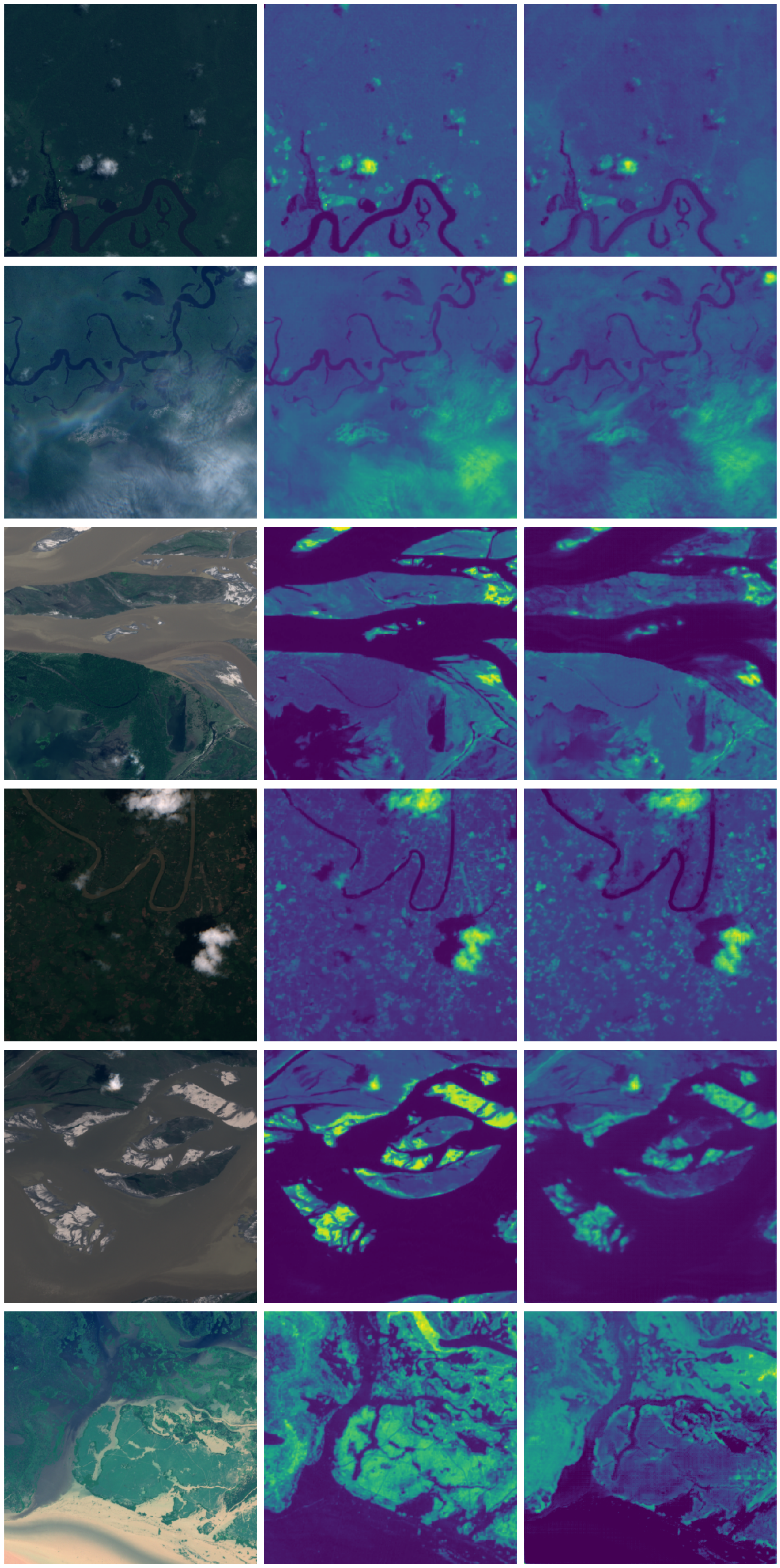

Figure 3: Qualitative results of SWIR-synth and ground truth $SWIR_2$ in low resolution imagery, Sentinel-2. Best viewed in color and zoomed.

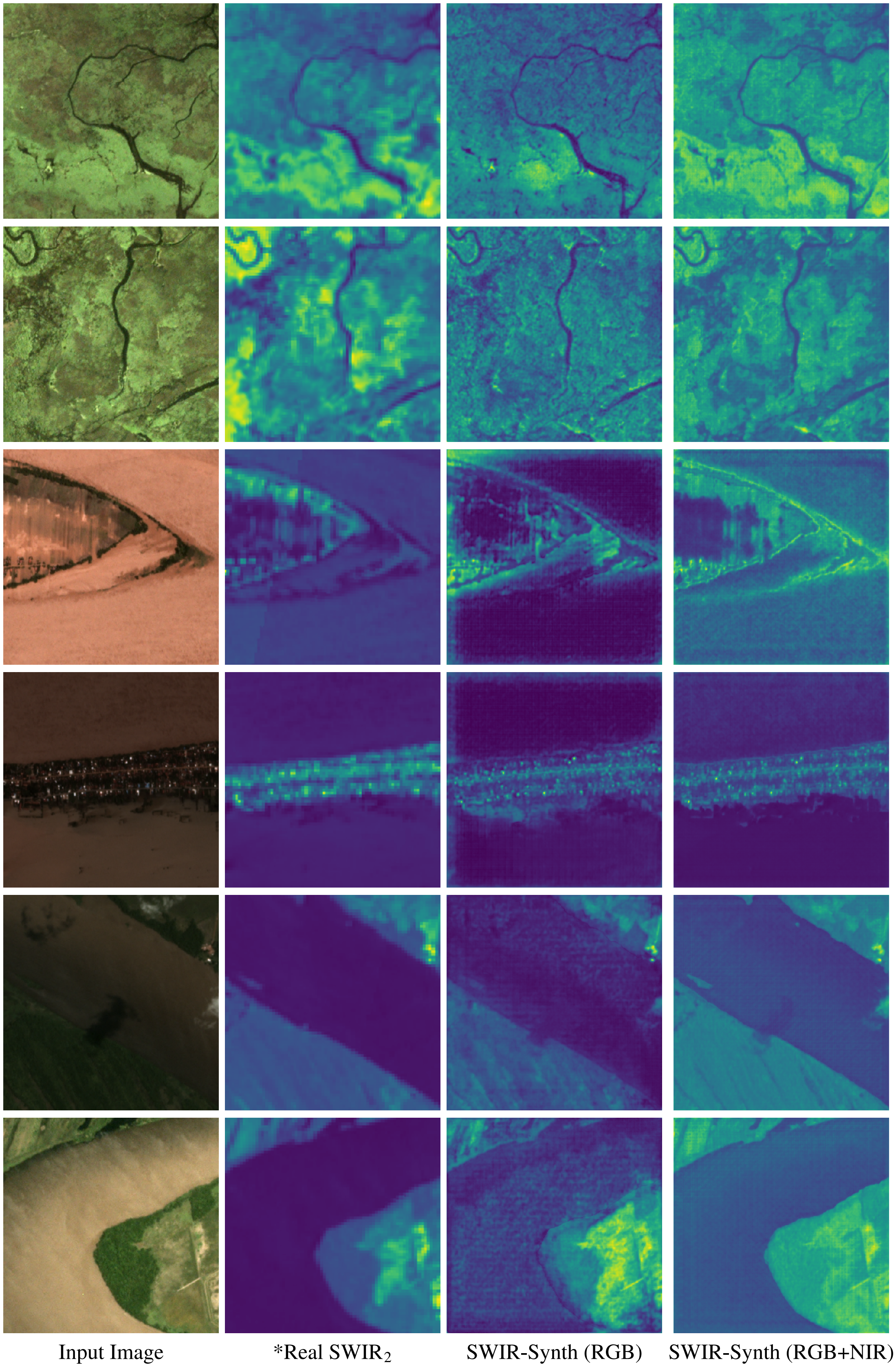

Figure 4: Qualitative results of SWIR-synth and *real $SWIR_2$ in high resolution imagery, PlanetScope. *Note that real $SWIR_2$ here is obtained through scarce overlapping crops with low resolution data. Further details is described in section 4.2 in the main paper. Best viewed in color and zoomed.

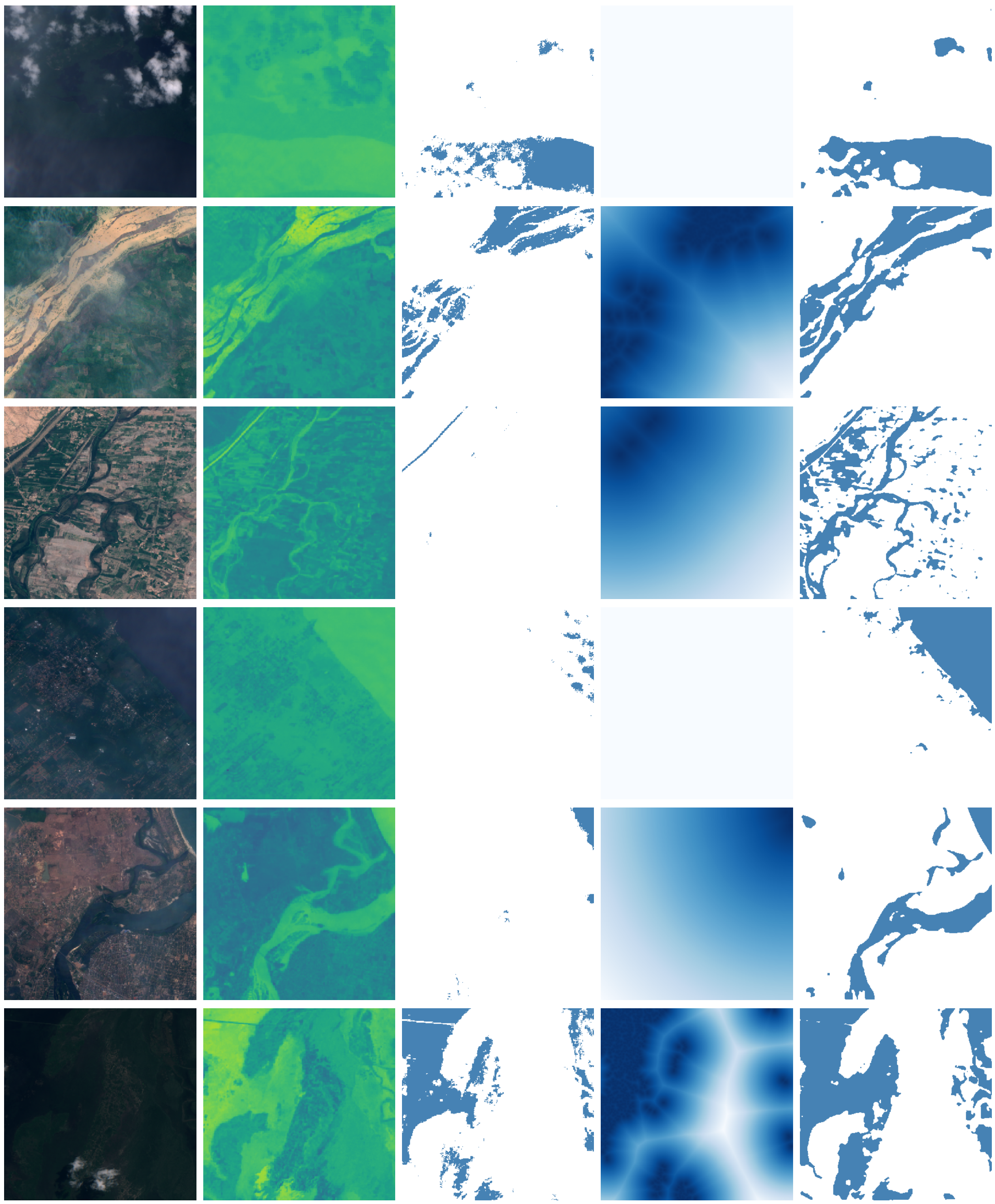

Figure 5: Qualitative results of the refiner on low resolution data from Sentinel 2 satellite. It can be seen that the refiner often completes hidden parts and refines overall mask. Best viewed in color and zoomed.

References

[1] Planetscope satellite imagery and archive, Sep 2020. 1, 2

[2] Dronedeploy. Dronedeploy segmentation benchmark dataset. 1

[3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1

[4] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *International Conference on Machine Learning*, pages 7354–7363, 2019. 1