Dual-Path Morph-UNet for Road and Building Segmentation from Satellite Images **Supplementary Material**

M. S. Dey, U. Chaudhuri, B. Banerjee, and A. Bhattacharya

S1 Backpropagation in Morphological Layers

To describe the backpropagation algorithm through the morphological layers, we start with a simple scenario. Consider a grayscale input image I of dimensions $A \times B$, which passes through a morphological layer to yield an output O. Let W denote the structuring element (SE) of the morphological layer, having a dimension of $L \times M$. We consider that the input is padded so that the input and output dimensions are the same. Let $\hat{O}$ denote the desired output of the operation. The objective of the network is to minimize the loss function $L = L(O, \hat{O})$, which represents the difference between O and $\hat{O}$. Given the value at location (l', m') of W, we calculate gradient using the chain rule as:

$$\begin{aligned}\frac{\partial L(O,\hat{O})}{\partial W(l',m')} &= \sum_{a=0}^{A-1}\sum_{b=0}^{B-1}\frac{\partial L}{\partial O(a,b)}\frac{\partial O(a,b)}{\partial W(l',m')} \\ &= \sum_{a=0}^{A-1}\sum_{b=0}^{B-1}\frac{\partial L}{\partial O(a,b)}\nabla O(a,b) \end{aligned} \tag{1}$$

Using the above equation, we can update the SE as:

$$W(l',m') = W(l',m') - \alpha\frac{\partial L}{\partial W(l',m')}$$

where α is the learning rate. The quantity that needs to be calculated for the gradients is ∇O. We now discuss the specifics for Dilation and Erosion.

Erosion. The output O from an Erosion layer is represented as

$$O = \min_{l,m}\{I(a+l, y+m) - W_e(l,m)\} \tag{2}$$

where W_e is the erosion SE. The $\min$ operation is not differentiable directly. However, for every $O(a,b)$, we can find a corresponding $W_e(l^*, m^*)$ which presents the best bet for the minimization problem. Thus Eq. 2 can be rewritten as

$$O(a,b) = I(a+l^*, y+m^*) - W_e(l^*, m^*)\}$$

Thus, ∇O can be expressed as:

$$\nabla O(a,b) = \begin{cases} -1 & \text{if } O(a,b) = I(a+l^*, y+m^*) - W_e(l^*, m^*)\} \\ 0 & \text{otherwise} \end{cases}$$

Dilation. Dilation is dual to the Erosion operation, and its the output O is expressed as

$$O = \max_{l,m}\{I(a-l, b-m) + W_d(l,m)\} \tag{3}$$

where W_d is the Dilation SE. Similar to Erosion, the $\max$ operation is non-differentiable directly. However for every $O(a,b)$ there exists a maximum $W_d(l^*, m^*)$, which is the best bet for the maximization problem. Thus, Eq. 3 can be expressed as

$$O(a,b) = I(a-l^*, b-m^*) + W_d(l^*, m^*)$$

such that ∇O can be expressed as

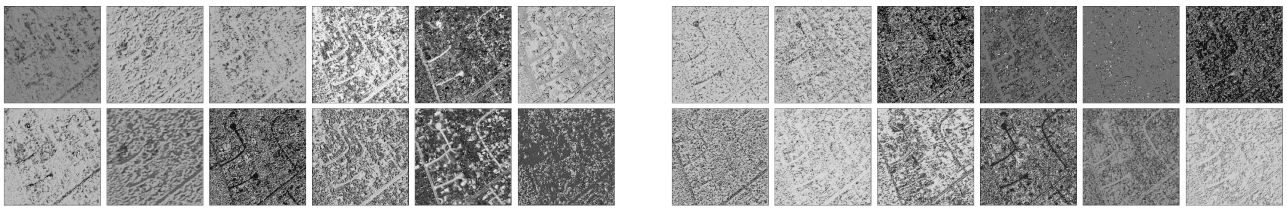
$$\nabla O(a,b) = \begin{cases} 1 & \text{if } O(a,b) = I(a-l^*, b-m^*) + W_d(l^*, m^*)\} \\ 0 & \text{otherwise} \end{cases}$$

The above treatment is similar to convolution backpropagation and can be easily extended to incorporate multiple Dilation and Erosion filters.

S2 Morphological Feature Maps

This section visualizes the feature maps obtained at each morphological block of the proposed architecture. We provide the visualizations for the Erosion filters as the Dilation filters predominantly enlarge each target region making the feature maps incomprehensible.

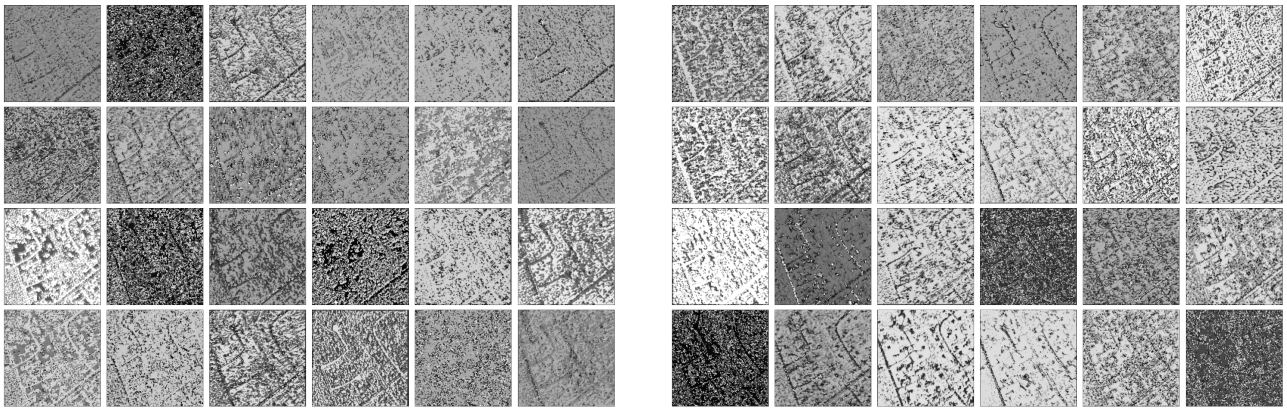
Before delving into the feature maps, we briefly describe the network architecture. The basic building block of the proposed network is the Morphological Unit (MU), comprising of learnable Dilation and Erosion filters, as illustrated in Fig. 1. of the manuscript. Six of these MUs are grouped similarly to the grouped convolutions, along with two BNConv blocks to form the Morphological Block (MB). As illustrated in Fig. 2 of the paper, the Dual-Path Block (DPB) utilizes dense and residual connections in sync with two MBs and a BNConv block to extract high-level morphological information. These DPBs are arranged in an encoder-decoder structure to form the proposed Dual-Path Morph-UNet. There are seven DPB in the network, with three as part of the encoder, one common block, and the last three as part of the decoder. To simplify the discussion, we denote the DPBs in encoder as EDPB_{12}, EDPB_{24} and EDPB_{36}, the common block as CDPB_{48} and those in decoder as DDPB_{36}, DDPB_{24}, DDPB_{12}. EDPB stands for Encoder Dual-Path Block, CDPB for Common Dual-Path Block, and DDPB denotes the Decoder Dual-Path Block. The subscript in the DPBs denotes the number of dilation and erosion feature maps generated by each of the encompassed MBs. The two MBs are represented as MB_1 and MB_2 for each of the DPBs. In the example below, we focus on an image patch belonging to the Massachusetts Road dataset of size 256×256 and observe the features selected by the network via Erosion filters. The feature maps from each MB are arranged in a grid of size $k \times 6$, where k is the number of channels of Erosion filters in an MU, and 6 denotes the number of MU in each MB.



(a) (b)

Figure S1: Erosion Maps belonging to (a) MB_1 and (b) MB_2 of $EDPB_{12}$

The output from the initial MU is sent to the first DPB in the encoder, which consists of 2 MBs. Each of the Erosion filters in an MU has a different set of SE, and we observe its effects on features selected as illustrated in Fig. S1. For example, in MB_1, the third MU extracts the outline of the road, while the fifth MU highlights the contour of the roads and the trees surrounding it. Similarly, in MB_2, the maps highlight the road outline, albeit very faintly in some cases. Compared to MB_1, the MB_2 tries to remove the background noise but gets overboard in some cases.

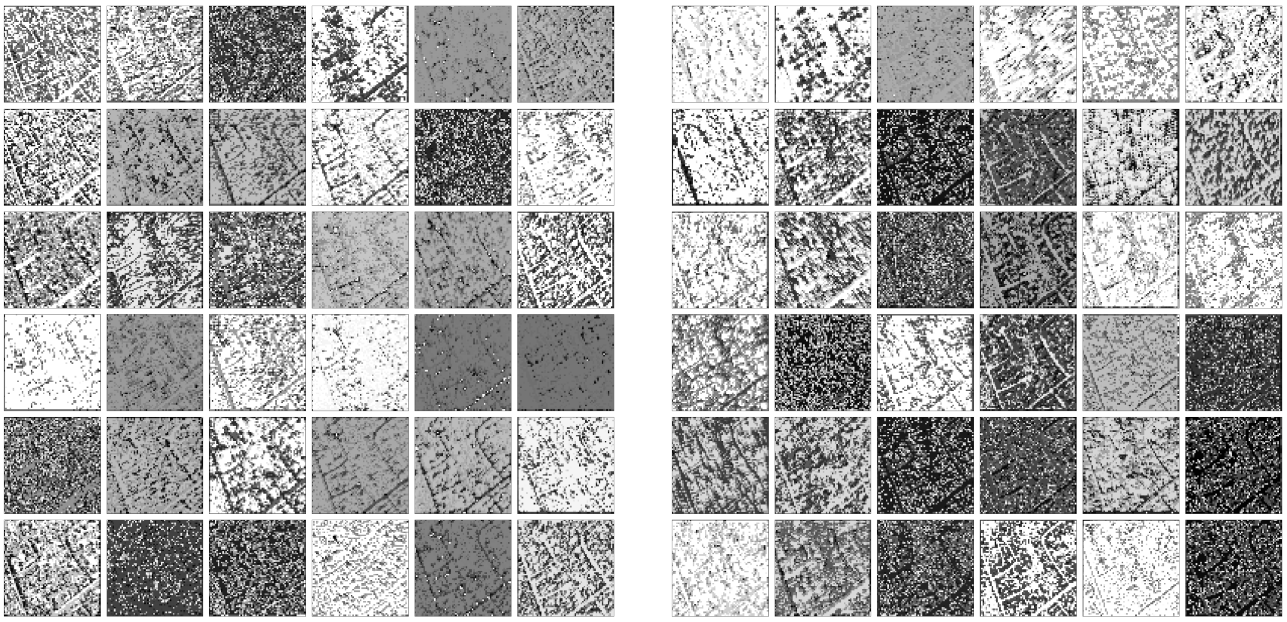


(a) (b)

Figure S2: Erosion Maps belonging to (a) MB_1 and (b) MB_2 of $EDPB_{24}$

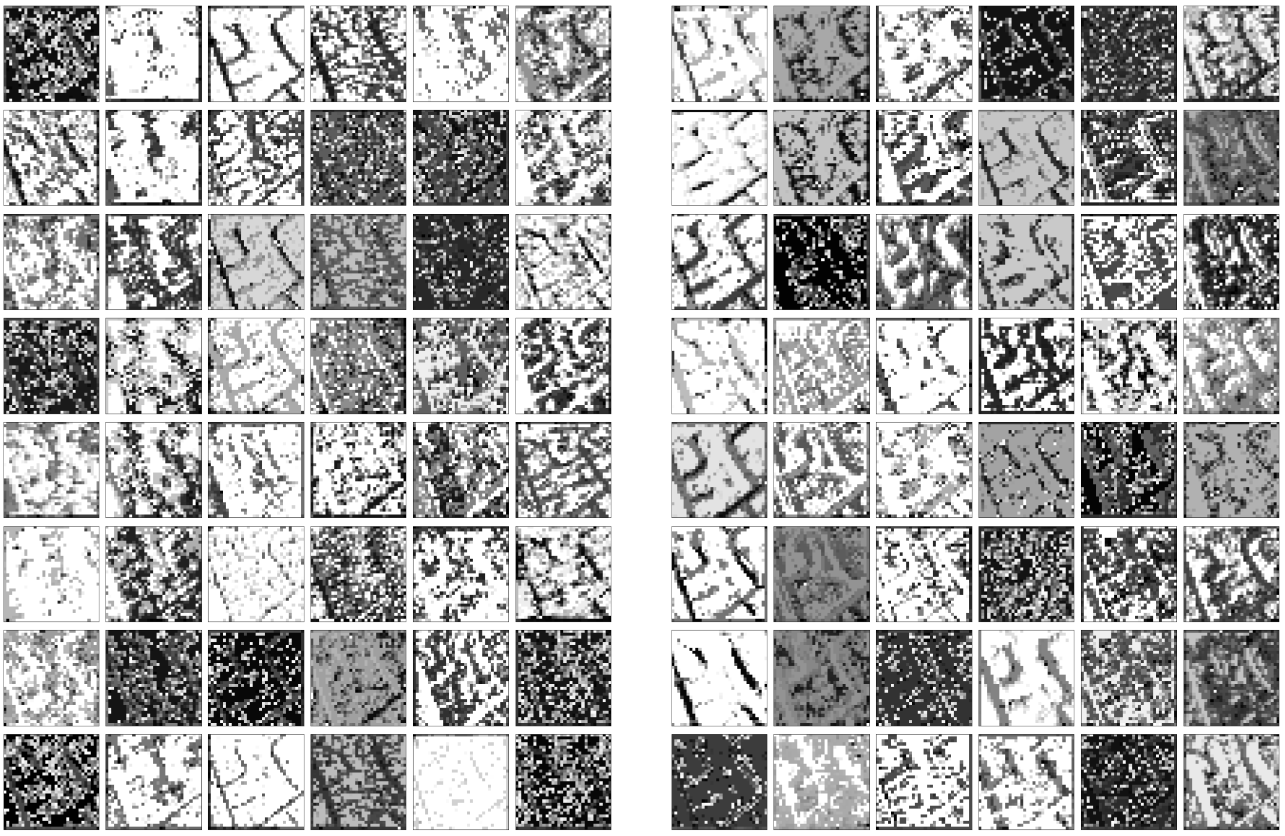
The output from the $EDPB_{12}$ is downsampled and passed onto $EDPB_{24}$. From Fig. S2 we observe that in MB_1 erosion filters start to highlight the road's outline. These features are further improved in MB_2, where the background noise is reduced while streamlining the road width.

The output from the $EDPB_{24}$ is downsampled and passed onto $EDPB_{36}$. By this stage, the network has a fair idea of the location of road segments in the image, as shown in Fig. S3. In both MB_1 and MB_2, it can be observed that the network tries to remove the background noise surrounding the road.



(a) (b)

Figure S3: Erosion Maps belonging to (a) MB_1 and (b) MB_2 of $EDPB_{36}$

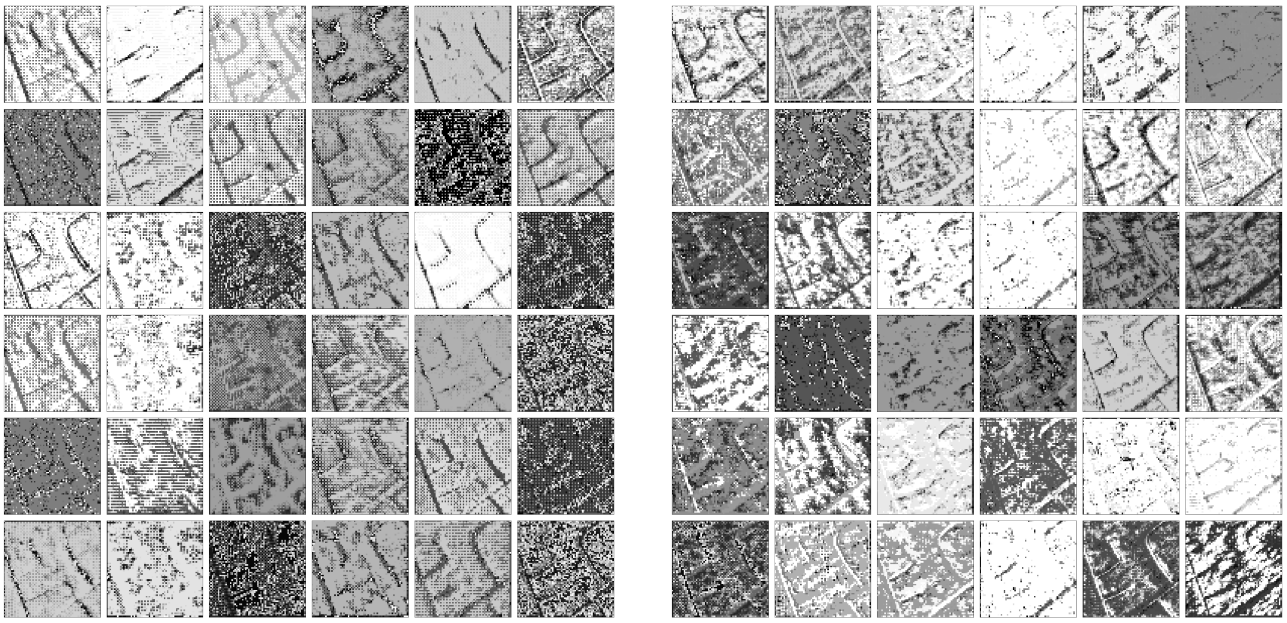


(a) (b)

Figure S4: Erosion Maps belonging to (a) MB_1 and (b) MB_2 of $CDPB_{48}$

The output from the $EDPB_{36}$ is downsampled and passed onto $CDPB_{48}$. As this point, due to the high dimensional data, some of the feature maps may appear to be incomprehensible, as shown in Fig S4. However, we observe that the network can now denoise the input of the background objects

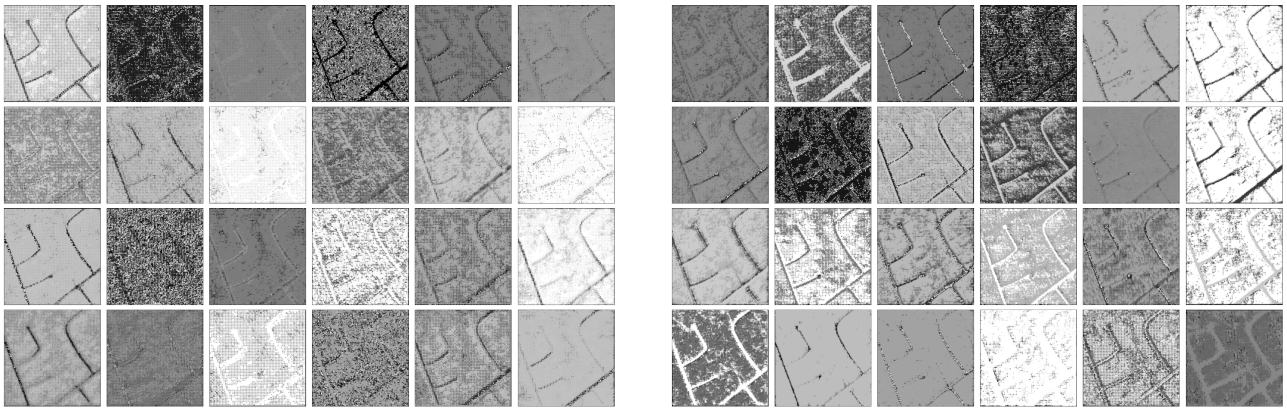
while getting the road outline. This can be prominently seen in the feature maps belonging to the third MU of MB_1, and the first and fourth MU of MB_2.



(a) (b)

Figure S5: Erosion Maps belonging to (a) MB_1 and (b) MB_2 of $DDPB_{36}$

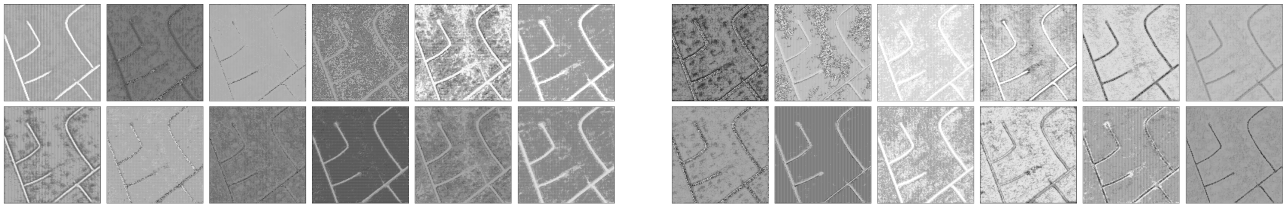
The output from the $CDPB_{48}$ is upsampled and passed onto $DDPB_{36}$. This is the first DPB in the Decoder block and is connected to the $EDPB_{36}$ in Encoder via skip connection. As seen in Fig. S5 compared to the previous blocks, the feature maps in MB_1 show a significant reduction in the noise near the roads, while in MB_2 there is a further smoothing of the input.



(a) (b)

Figure S6: Erosion Maps belonging to (a) MB_1 and (b) MB_2 of $DDPB_{24}$

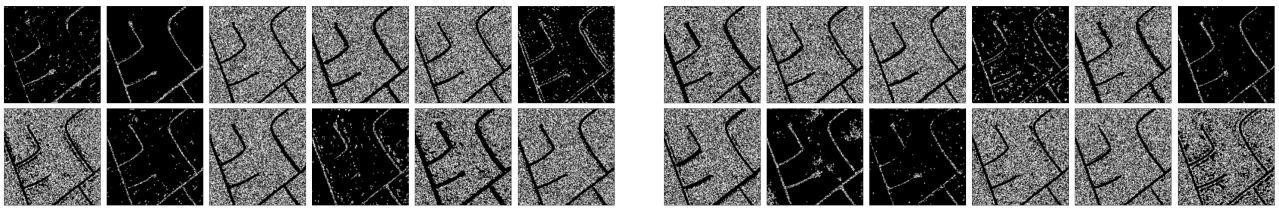
The output from the $DDPB_{36}$ is upsampled and passed onto $DDPB_{24}$. By this point, the network seems to locate the road while minimising the background objects, as observed from Fig S6. The MB_2 further reduces the noise while prominently highlighting the road.



(a) (b)

Figure S7: Erosion Maps belonging to (a) MB_1 and (b) MB_2 of $DDPB_{12}$

The output from the $DDPB_{24}$ is upsampled and passed onto $DDPB_{12}$. As can be seen from Fig S7, in both MB_1 and MB_2, the network delineates the road while denoising the background. Unlike previous blocks, we also observe that the road contour and thickness closely resembles the ground truth. This is the last block of the Decoder, and its output is passed to the sigmoid layer to get the final segmented result.

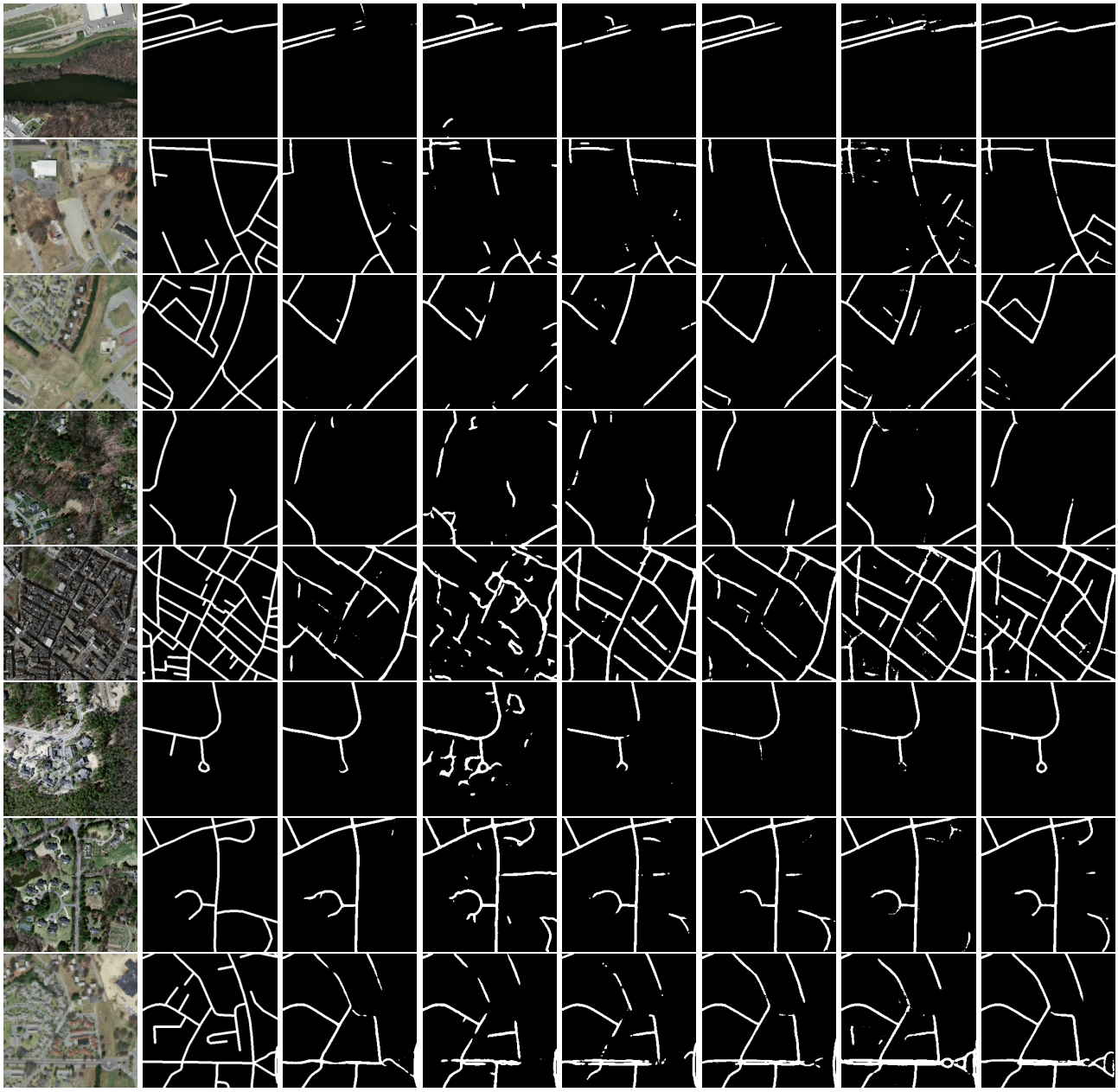


(a) (b)

Figure S8: Pre final layer feature maps of Conv. UNet

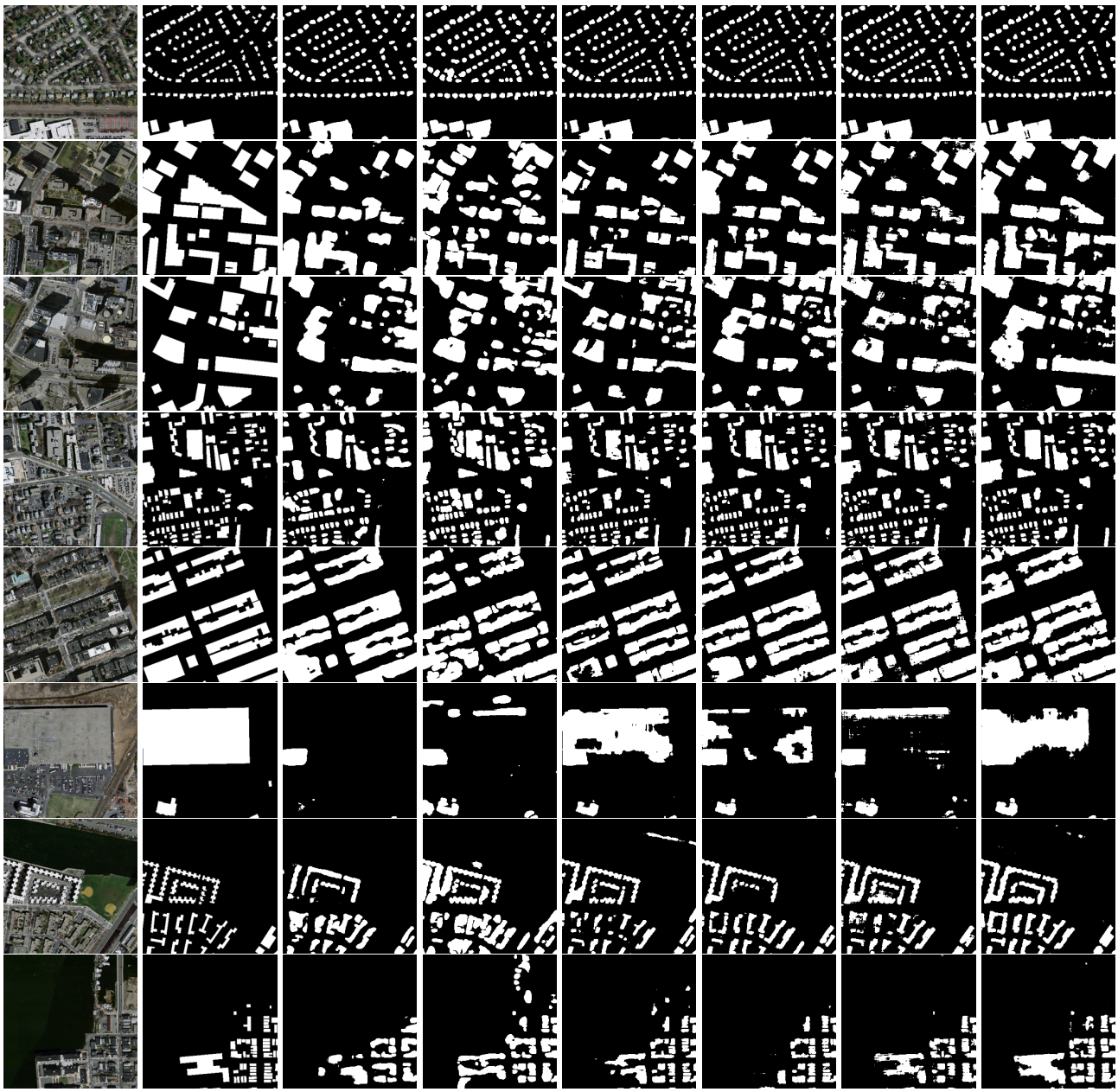
We also compare the pre-final layer feature maps from the proposed network (Fig S7) with the one obtained from Conv. UNet, as illustrated in Fig. S8. While both the feature maps can localize the road segments, there are a couple of subtle yet important differences. In many of the convolution feature maps, the roads segments seem disconnected from each other, while in some there is noticeable background noise. Furthermore, the contours and the thickness of the road segments are not uniform. By comparison, in almost all the erosion feature maps, the road segments are connected and exhibit uniform contours while background noise is significantly reduced. This clearly demonstrates the critical role morphological features play in the R.S. object segmentation, and the efficacy of the proposed Dual Path Morph-UNet in incorporating the same.

S3 Segmented Output



(a) (b) (c) (d) (e) (f) (g) (h)

Figure S9: (a) Satellite image, (b) ground truth, and segmentation results of (c) LinkNet (d) UNet++ (e) Residual UNet (f) Attention UNet (g) JointNet (h) Dual Path Morph-UNet on the Massachusetts Roads dataset.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure S10: (a) Satellite image, (b) ground truth, and segmentation results of (c) LinkNet (d) UNet++ (e) Residual UNet (f) Attention UNet (g) JointNet (h) Dual Path Morph-UNet on the Massachusetts Building dataset.