VStegNET: Supplementary Material

Contents

List of Figures

References

[1] Shumeet Baluja. "Hiding Images in Plain Sight: Deep Steganography". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 2069–2079. URL: http://papers.nips.cc/paper/6802-hiding-images-in-plain-sight-deep-steganography.pdf.

Figure 1: ($\beta = 0.75$), Left to right: cover (C), container (C'), ($C - C'$), secret (S), revealed secret (S'), ($S - S'$)

Figure 2: ($\beta = 1.0$), Left to right: cover (C), container (C'), ($C - C'$), secret (S), revealed secret (S'), ($S - S'$)

Figure 3: [1], Left to right: cover (C), container (C'), ($C - C'$), secret (S), revealed secret (S'), ($S - S'$)

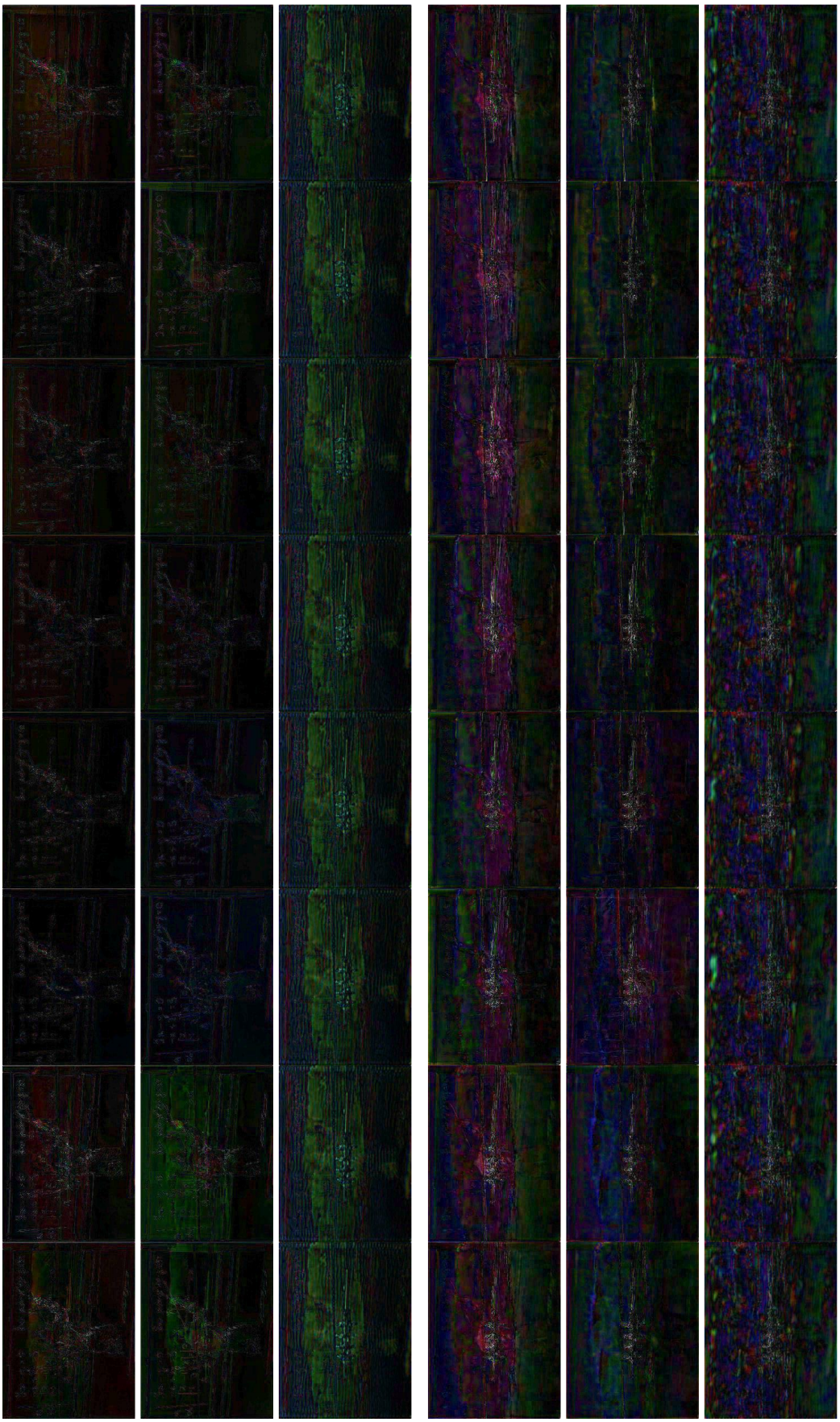

Figure 4: Residual $\times 5$, Left to right: $[C - C']$ - $(\beta = 0.75)$, $(\beta = 1.0)$, [1], $[S - S']$ - $(\beta = 0.75)$, $(\beta = 1.0)$, [1]

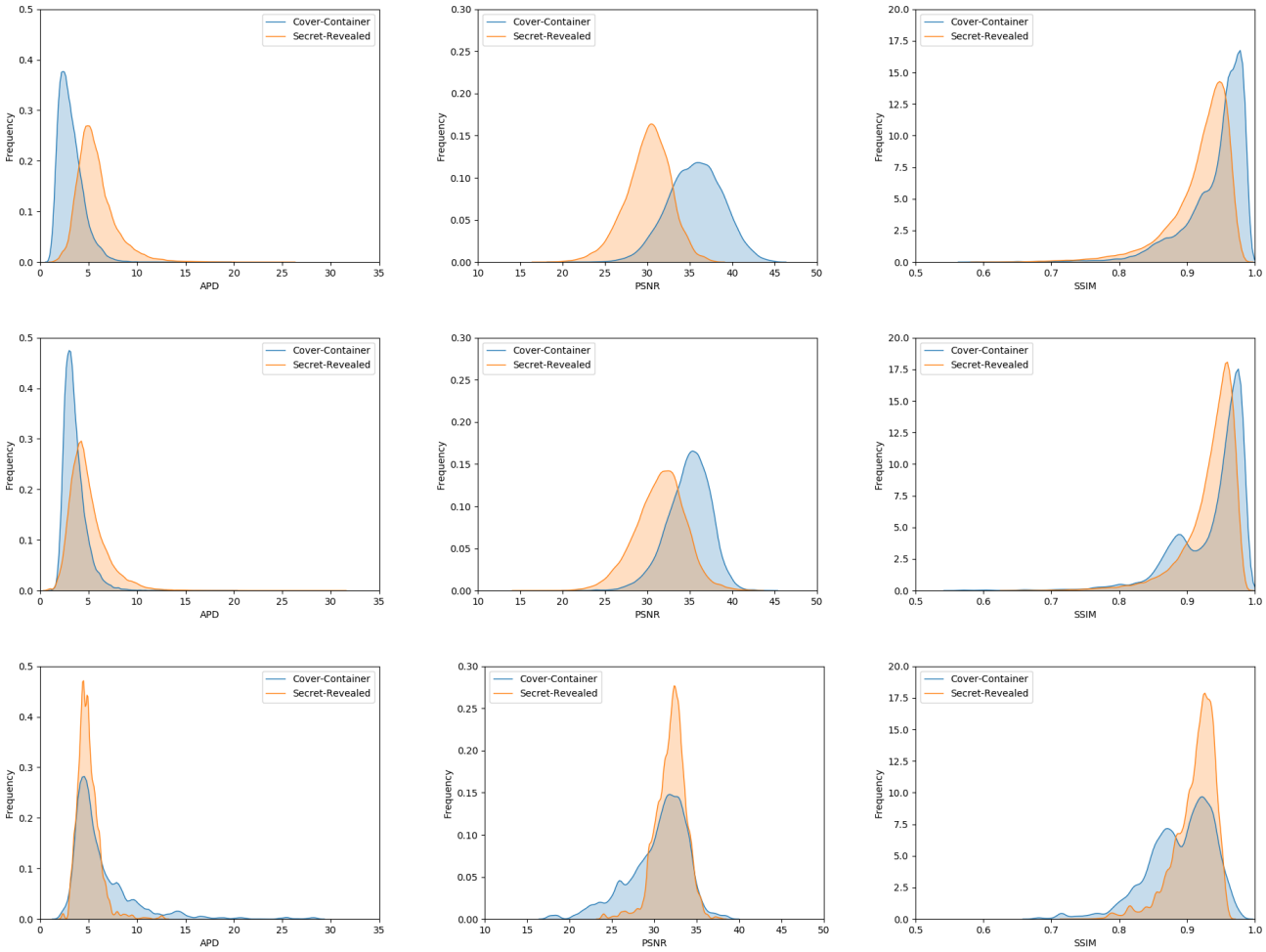

Figure 5: Histogram of APD, PSNR and SSIM values on the test dataset. First row: ($\beta = 0.75$), Second row: ($\beta = 1.0$), Third row: [1]. One can see that our models outperform [1] on each of these metrics.

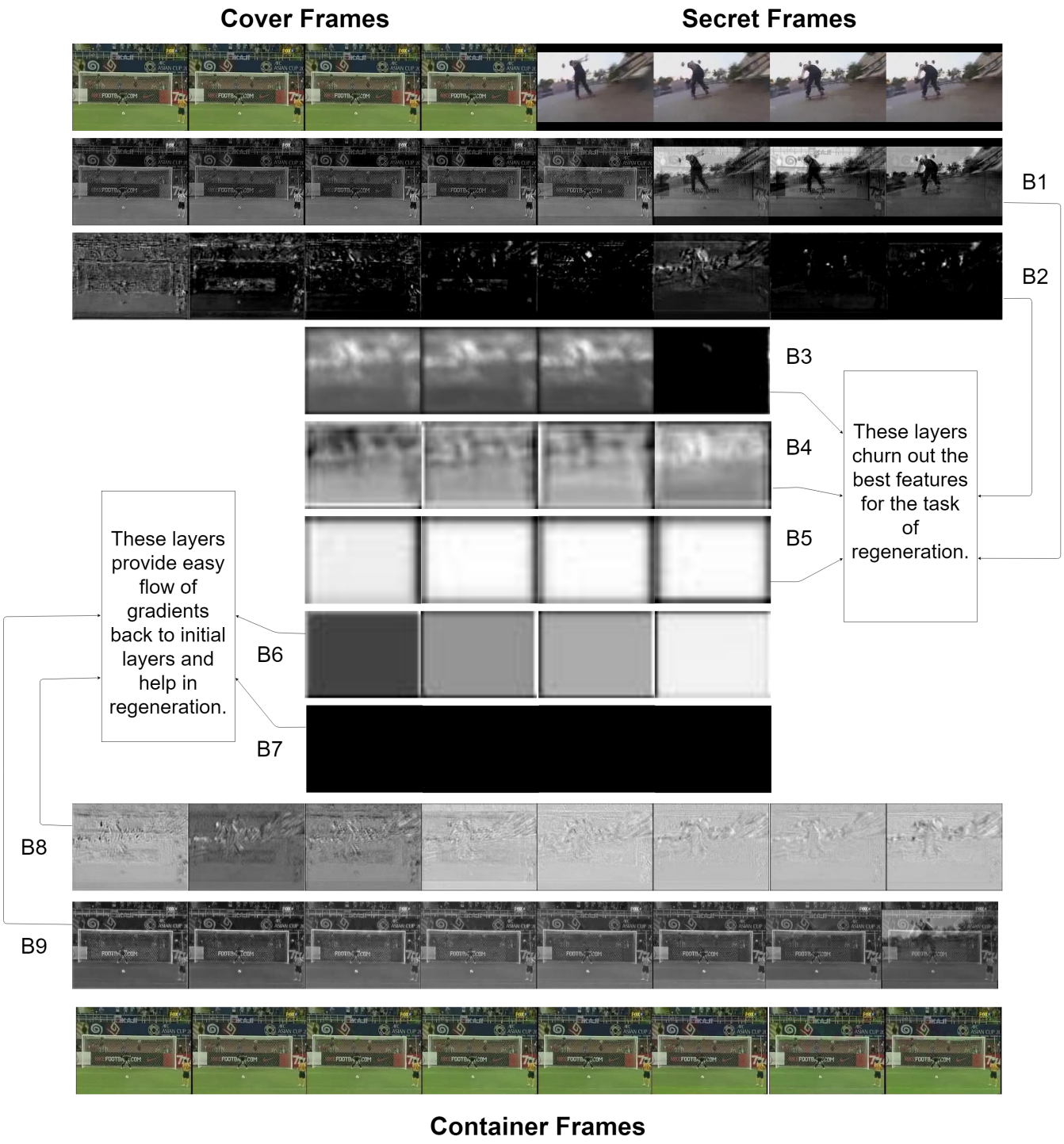

Figure 6: Activation maps produced at the end of each block of the VHN. Each row shows the activation produced for corresponding frames above, by a sample filter from previous layer. One can see the gradual hiding of secret frames inside cover frames (B1). After B2, features have been encoded and max-pooled in the temporal dimension as well, reducing the number of frames to half. Features are extracted in a form that is visually not perceivable. After the bottleneck, layers have skip connections for better performance (See Section 3 in the paper). After B7, the decoder part of the network starts producing visually discernible container frames. The container frames showed here are the total resultant frames from 16 input frames (8 middle frames are shown in the first row.)

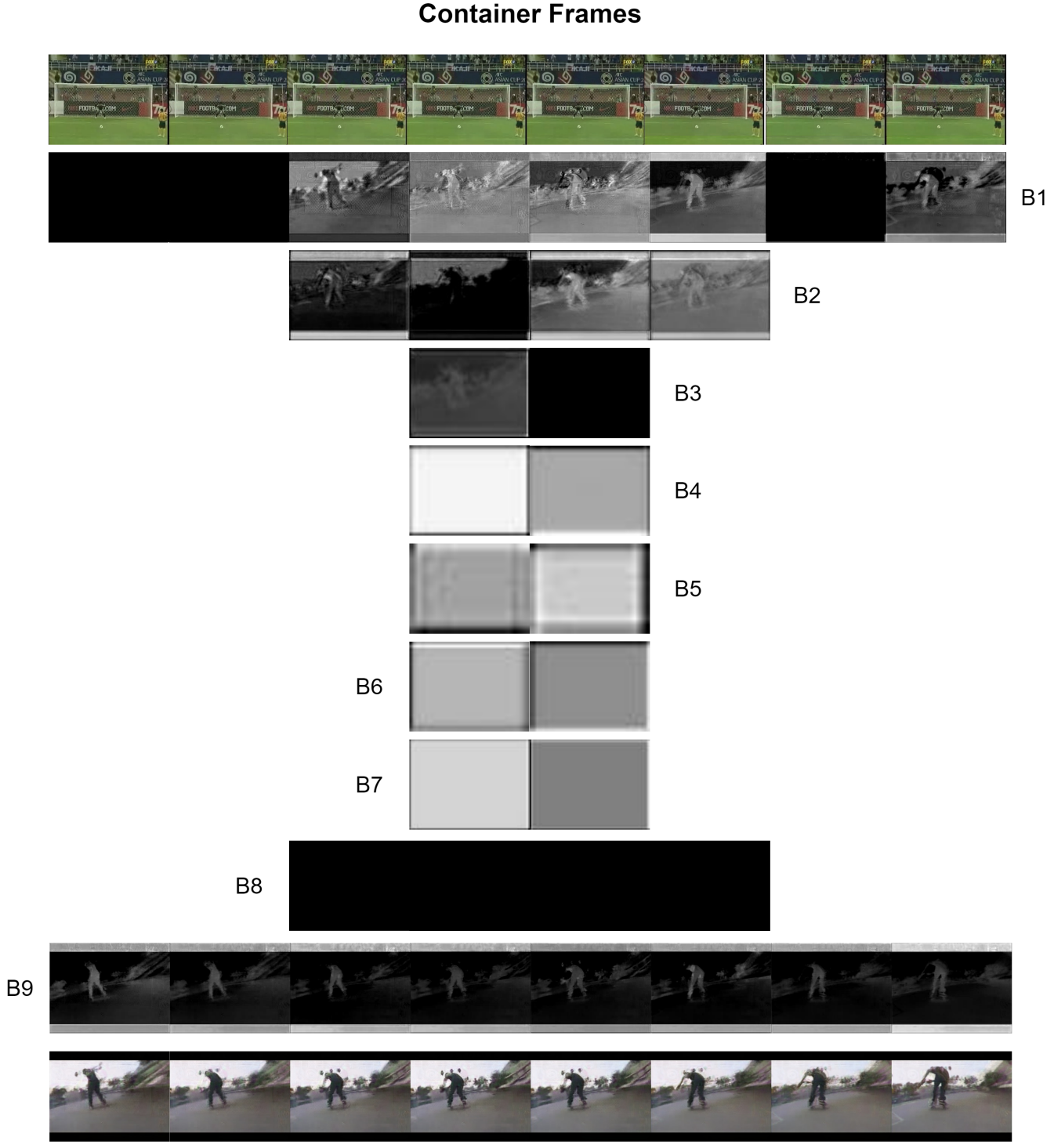

Figure 7: Similar to the VHN, the activations after each block of the VRN are shown row-wise. Different types of activations are produced (B1) from visually similar container frames, using the same filter. This is because each filter is of a small shape $(3 \times 3 \times 3)$. This enables the VHN to hide secret frames in very small portions of the cover frames that are not visible to the naked eye, but these small filters are able to bring out that information. Some of the activations are blank because each filter sees something different in a particular image. Similar to VHN, skip connections help produce the secret frames back, after the bottleneck.