DOA-GAN: Dual-Order Attentive Generative Adversarial Network for Image Copy-move Forgery Detection and Localization

Ashraful Islam[1], Chengjiang Long[2]*, Arslan Basharat[2], and Anthony Hoogs[2]
[1]Rensselaer Polytechnic Institute, Troy, NY
[2]Kitware Inc., Clifton Park, NY
islama6@rpi.edu, {chengjiang.long, arslan.basharat, anthony.hoogs}@kitware.com

Abstract

The supplementary material provides additional visualization results on the three benchmark datasets, invariance analysis and details of the extension of DOA-GAN to image splicing and video copy-move forgery.

1. Invariance Analysis

In the main paper, we performed robustness analysis of DOA-GAN under different attacks for COMO dataset. Here we provide analysis on our self-collected dataset generated from MS COCO. Particularly, we created our CMFD dataset from MS COCO following similar approach to the USC-ISI CMFD dataset [7], and divided the dataset into six groups, based on the type of transformations applied in the spliced regions, namely, raw (no transformation), scale, rotation, blur, flip and luminance. Each group consists of $5,000$ CMFD images.

For more specific detail of the above transformations, the scale ratio is in a range of $[0.75, 3.0]$, the rotation angle in a range of $[-45°, 45°]$, the kernel size for average blurring involves 2×2 to 5×5, the parameter of contrast normalization for luminance is from 0.75 to 1.5, and either vertical or horizontal flip is applied randomly.

Figure 1 shows the performance for different kind of transformations applied for both source and target/forged masks, in terms of F1 score. It shows that the localization scores for source and forged mask on blurring, scaling, and luminance are much higher, hence the model is quite robust to those transformations. Though the score declines on rotation and flip, it is still around 40% for source mask and 60% for target mask. Note that we get the best score for target mask on the Blur set, as blurring distorts the target region in such a way that it is easily distinguishable.

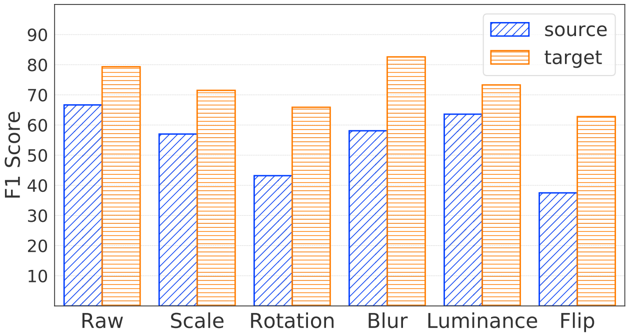

Figure 1: Invariance Analysis on our self-collected dataset generated from MS-COCO.

2. Experiments

To verify the effectiveness of our proposed DOA-GAN for copy-move forgery detection and localization, we conduct experiments on three benchmark datasets: the USC-ISI CMFD dataset [7], the CASIA CMFD dataset [7], and the CoMoFoD dataset [5].

2.1. Experiments on the USC-ISI CMFD dataset.

To further understand the advantage of our proposed dual-order attention GAN, we also provide additional visualization results in Figure 2. As we can see, our DOA-GAN is able to generate more accurate masks than BusterNet, our FOA-GAN (First-Order Attention GAN), and our SOA-GAN (Second-Order Attention GAN).

2.2. Experiments on the CASIA CMFD dataset.

In Figure 3, we provide visualization results on some examples of CASIA CMFD dataset. We see that our proposed DOA-GAN is able to detect more accurate masks than DenseField and BusterNet for the copy-move forgery manipulation, although it is less accurate than the groundtruth masks.

*This work was supervised by Chengjiang Long when Ashraful Islam was a summer intern at Kitware Inc.

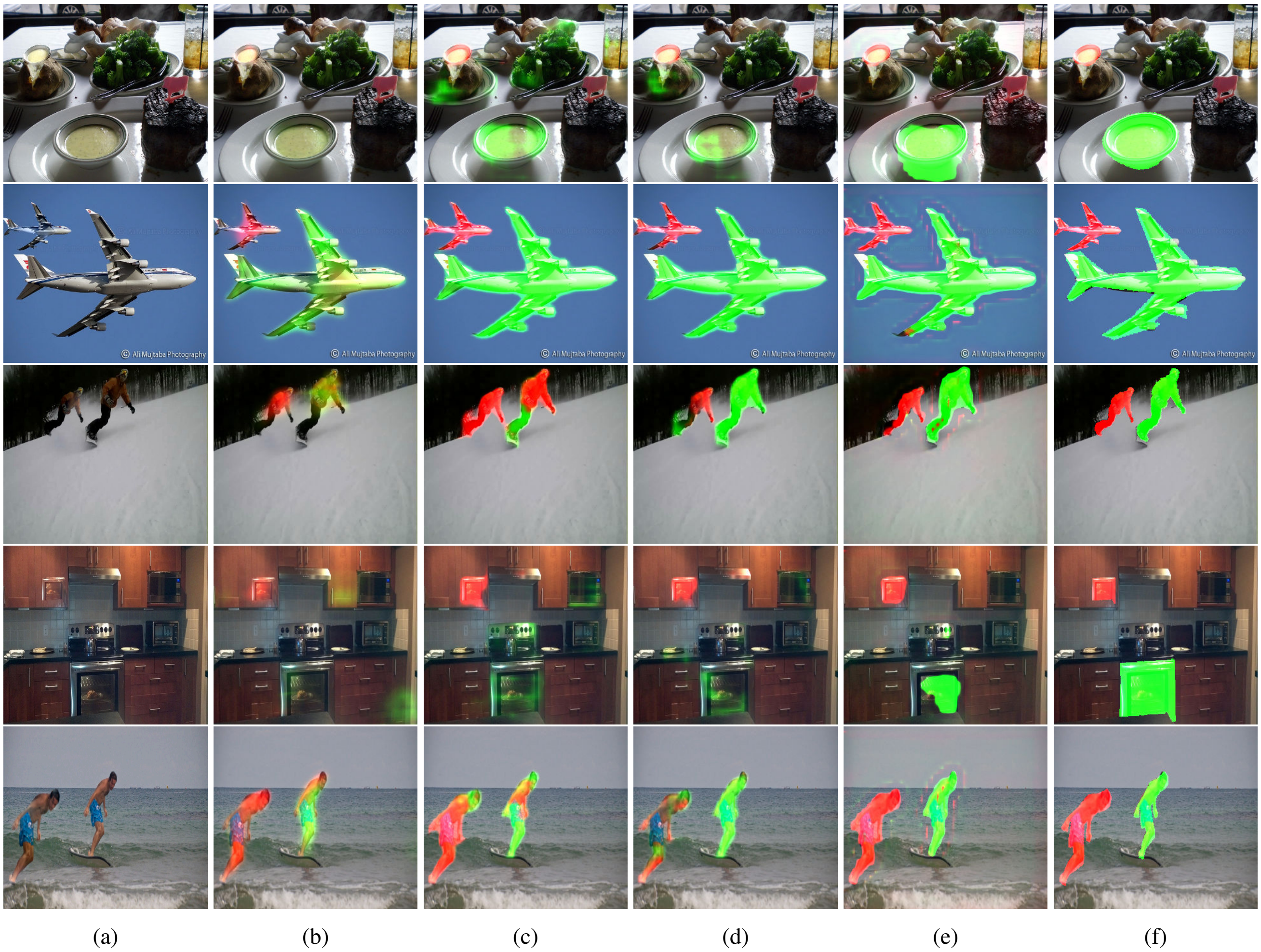
(a) (b) (c) (d) (e) (f)

Figure 2: Some qualitative results on USC-ISI CMFD dataset. From left to right are the input images (a), results of BursterNet [7] (b), results of our FOA-GAN (c), results of our SOA-GAN (d), results of our DOA-GAN (e), and the ground truth masks (f).

2.3. Experiments on the CoMoFoD dataset.

We provide some examples in Figure 4 for the visualization comparison on CoMoFoD dataset.

3. Extension of DOA-GAN for Image Splicing Detection and Localization

The proposed model can be extended to image splicing detection and localization with minor modifications in the network architecture. In particular, given two images, one of which is spliced image and the other one is donor image, we calculate affinity matrix on the extracted features obtained from two separate feature extractor modules - one for probe image and another for donor image. From the affinity matrix and contextual features from ASPP blocks, we obtain first-order and second-order attention features, and merge them to obtain two final feature representations, which are fed into two separate convolution blocks to predict source mask and target/forged mask.

Note that we do not use the Gaussian operator G in the attention module for image splicing. Our model is trained on a synthetic image splicing dataset, consisting of $87K$ training image pairs, following the generation process described in [2]. Figure 5 shows the full framework. We also show some visualization comparisons in Figure 6.

4. Extension of DOA-GAN for Video Copy-move Forgery

Video Copy-Move Forgery denotes copying a compact video object and inserting into different location, either in the same frames or different frames. We formulate video CMFD as an extension to image splicing detection and localization. We propose a video copy-move forgery de-

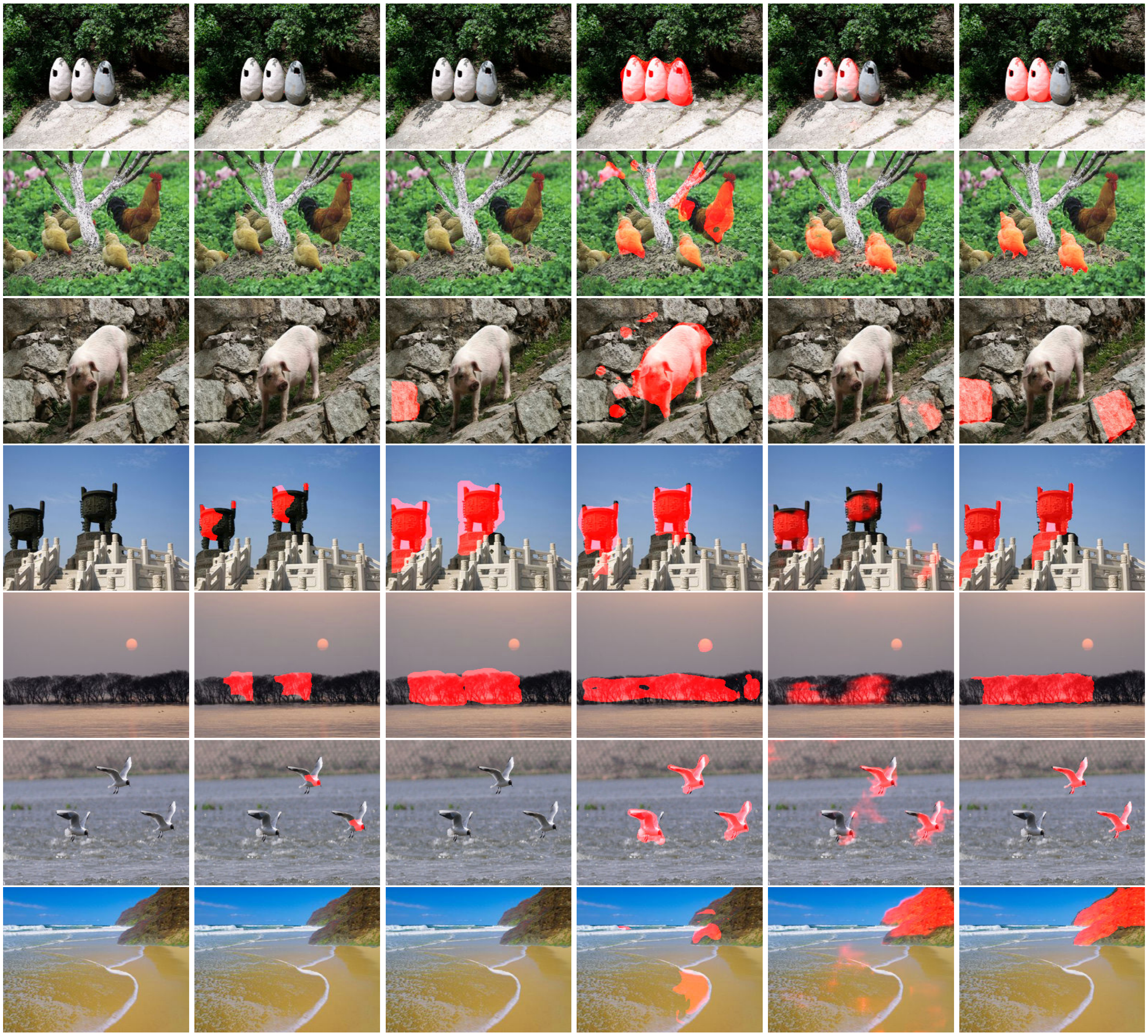

Figure 3: Visualization examples on the CASIA CMFD dataset. From left to right are the input image; results of Adaptive-Seg, DenseField, BursterNet, and our DOA-GAN; and the ground truth mask.

tection algorithm that utilizes the image splicing detection framework to generate inter-frame masks and construct confusion matrix, and predicts the final output mask based on temporal consistency. In particular, given a video, we first determine a set of candidate image pairs, and generate source and target/forged masks by an inter-frame splice detection framework. The confusion matrix is constructed from the localization mask of each image pair, and contains the probability score that the image pair has spliced forgery. After that, we calculate the most probable continuous frames of source and spliced regions by finding a line parallel to the diagonal line of the confusion matrix based on the confusion scores. Note that our proposed DOA-GAN is used for inter-frame splice detection and localization in this framework. It is worth mentioning that in this paper we currently take all the possible N^2 image pairs because of short videos used in our experiments, where N is the number of video frames, and we plan to deal with long videos in the near future.

For lack of video CMFD dataset, we generate a synthetic dataset from video object segmentation datasets, namely, DAVIS [3], SegTrackV2 [6] and Youtube-object [4]. Our generated dataset consists of 240 training videos and 120 testing videos. We compare our method with DMAC, DMVN, and PatchMatch3D [1]. Note that DOA-GAN, DMAC and DMVN are used for inter-frame splicing de-

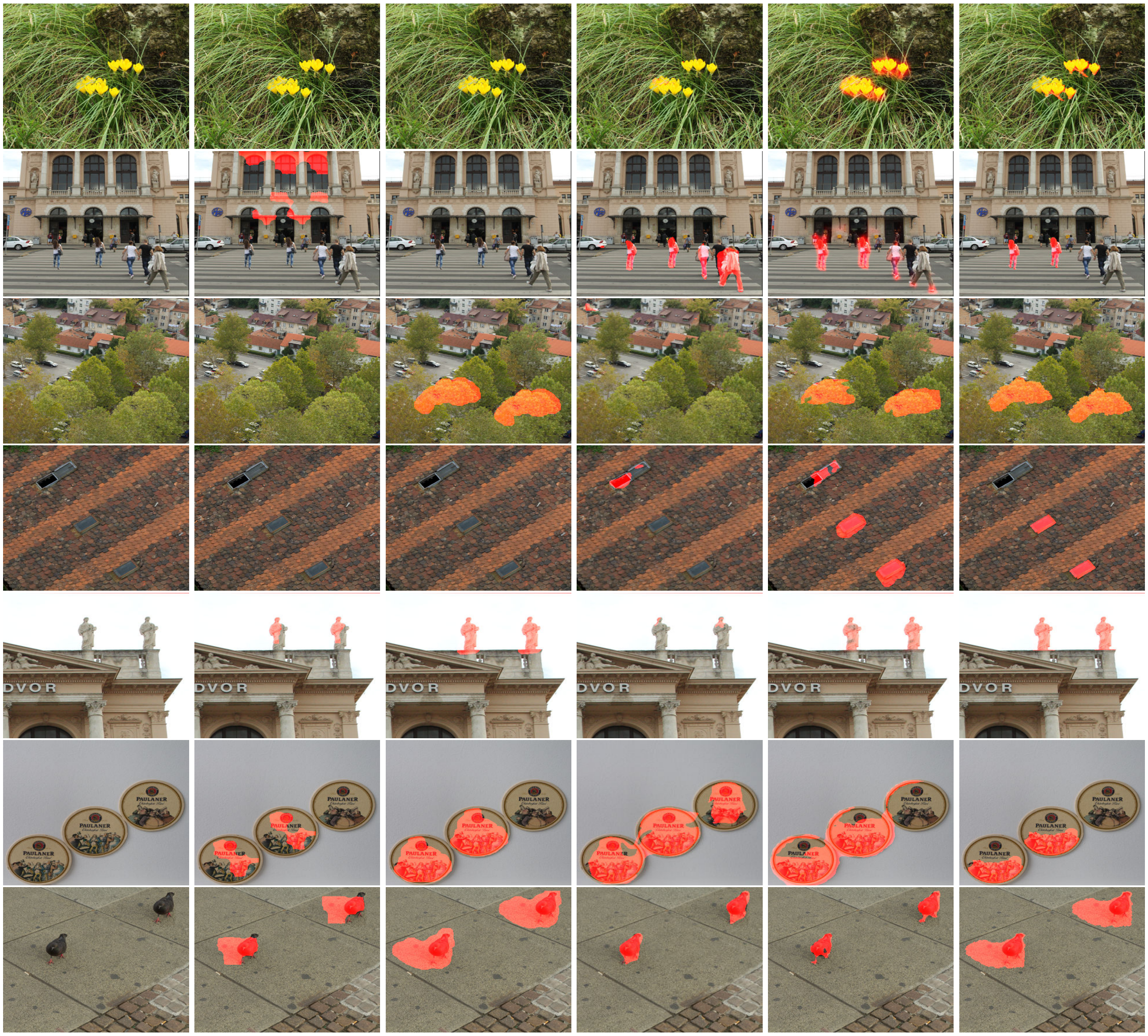

Figure 4: Visualization examples on the CoMoFoD dataset. From left to right are the input image; results of Adaptive-Seg, DenseField, BursterNet, and our DOA-GAN; and the ground truth mask, respectively.

tection and localization. We provide a visualization examples in Figure 7, from which we can see that our model can clearly distinguish source and target regions.

4.1. Discussion

Through the above experiments, we demonstrated the promising advantages of our proposed DOA-GAN, which is able to use the copy-move region attention to extract manipulation attentive features, as well as the co-occurrence feature with patch-to-patch interdependence taken into consideration. Regarding the running time, it takes 0.070 seconds to process an image of size 320 × 320.

Acknowledgement

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA) under Contract No.FA875016-C-0166. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

References

[1] Luca D'Amiano, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. A patchmatch-based dense-field

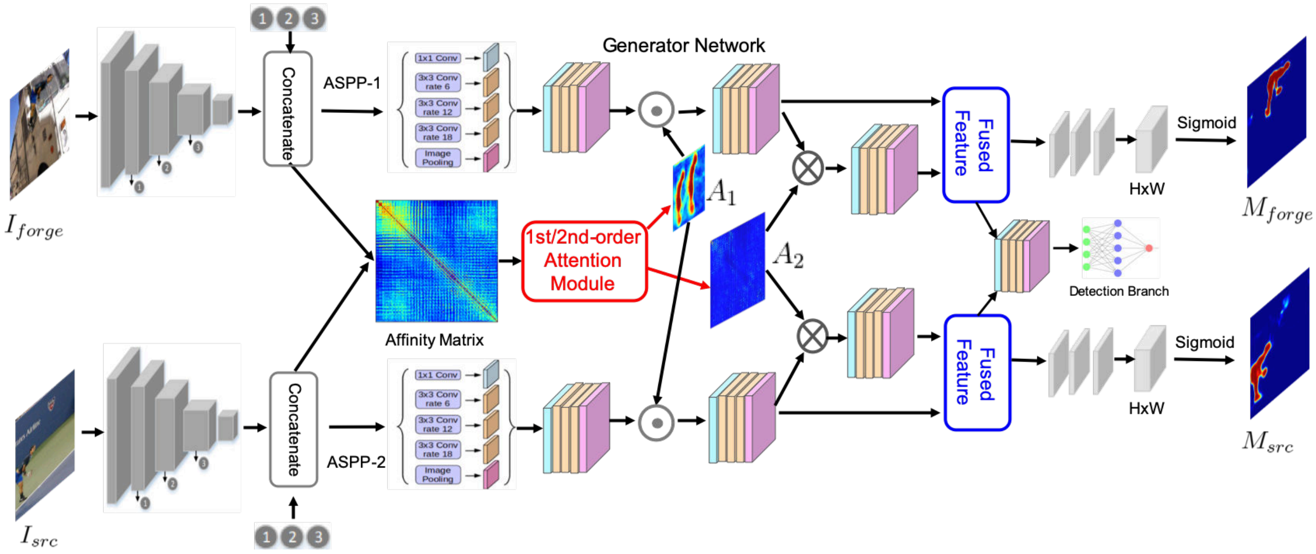

Figure 5: The overview of the modified DOA-GAN framework for image splicing forgery detection and localization. Note that we do not show the discriminator branch for simplicity.

algorithm for video copy–move detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(3):669–682, 2018.

[2] Yaqi Liu, Xianfeng Zhao, Xiaobin Zhu, and Yun Cao. Adversarial learning for image forensics deep matching with atrous convolution. *arXiv preprint arXiv:1809.02791*, 2018.

[3] Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 724–732, 2016.

[4] Alessandro Prest, Christian Leistner, Javier Civera, Cordelia Schmid, and Vittorio Ferrari. Learning object class detectors from weakly annotated video. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3282–3289. IEEE, 2012.

[5] Dijana Tralic, Ivan Zupancic, Sonja Grgic, and Mislav Grgic. CoMoFoD–new database for copy-move forgery detection. In *ELMAR*, pages 49–54, 2013.

[6] David Tsai, Matthew Flagg, and James M.Rehg. Motion coherent tracking with multi-label mrf optimization. *BMVC*, 2010.

[7] Yue Wu, Wael Abd-Almageed, and Prem Natarajan. BusterNet: Detecting copy-move image forgery with source/target localization. In *ECCV*, pages 168–184, 2018.

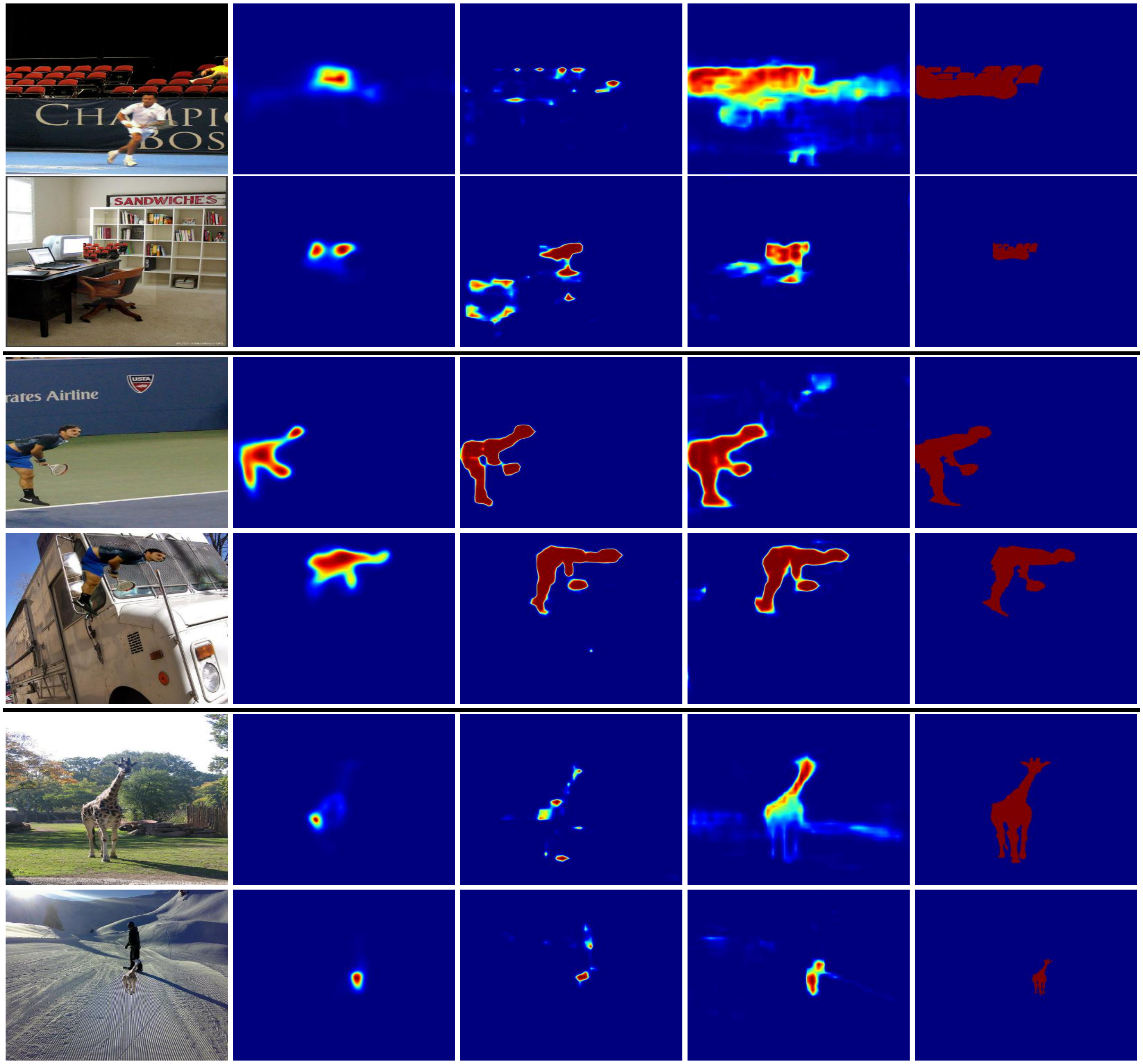

Figure 6: Visualization examples on generated splicing dataset for image splicing localization. From left to right are the input images, results of DMVN, DMAC, DOA-GAN, and ground-truth mask, respectively.

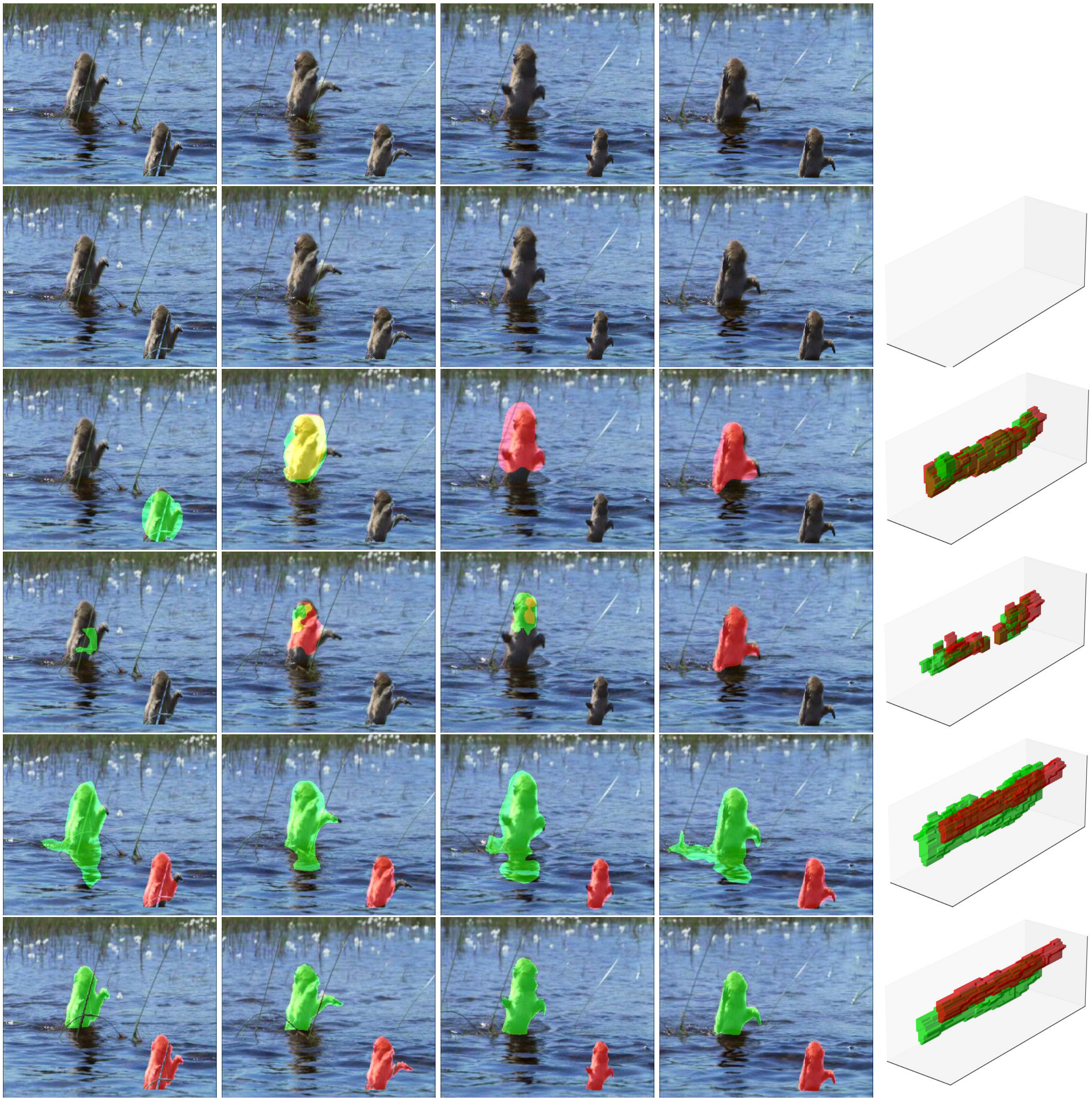

Figure 7: Visualization of video copy-move localization. From top to bottom are frames sampled form an input video, 2D image region masks and 3D copy-move masks (rightmost column) for 3D Patch-Match, DMVN, DMAC, our DOA-GAN, and ground-truth. Here, source mask is annotated by green and forge mask is annotated by red. In the 3D view, the time axis is on the bottom right. The rightmost column represents a 3D view, where the time axis is on the bottom right.