Blind Visual Motif Removal from a Single Image

Amir Hertz Sharon Fogel Rana Hanocka Raja Giryes Daniel Cohen-Or
Tel Aviv University

Figure 1: Blind visual motif removal results on images unseen during training. Top: test images embedded with semi-transparent motifs. Bottom: our reconstructed results. Our network was trained on Latin characters, yet successfully identifies and removes the Hindi and Japanese characters (left three images). Similarly, the overlaid visual motifs on the right three images differ semantically from the motifs used during training.

1. Overview

Many images shared over the web include overlaid objects, or *visual motifs*, such as text, symbols or drawings, which add a description or decoration to the image. For example, decorative text that specifies where the image was taken, repeatedly appears across a variety of different images. Often, the reoccurring visual motif, is semantically similar, yet, differs in location, style and content (*e.g.,* text placement, font and letters). This work proposes a deep learning based technique for *blind* removal of such objects. In the blind setting, the location and exact geometry of the motif are unknown. Our approach simultaneously estimates which pixels contain the visual motif, and synthesizes the underlying latent image. It is applied to a single input image, without any user assistance in specifying the location of the motif, achieving state-of-the-art results for blind removal of both opaque and semi-transparent visual motifs.

The removal of these visual motifs and the recovery of a pristine image can be an extremely challenging task. The structure, size and location of these objects varies between different images, making them difficult to detect without user guidance or assumptions about the underlying image. Previous methods have relied on information about the location of the corrupted pixels to be restored [4, 3, 5, 9]. Dekel et al. [1] remove watermarks using large image collections, which contain the same watermark, as well as some minimal user guidance about the watermark location.

We present a method for completely *blind* visual motif removal. In the blind setting, the exact location, structure and size of these motifs is unknown. The generalization ability of our network is demonstrated by removing visual motifs that are not seen during training (See Figure 1), and naturally, our generalization can be *abused* by removing visual watermarks from protected images. See Figure 2 for examples of removing watermarks from various stock photography services. Unlike previous approaches, our strategy does not require multiple images with the same object to be removed, or the exact location of the motif pixels.

2. Method

Our proposed approach tackles this problem using a convolutional neural network (CNN) trained to remove visual motifs embedded in an image. We train the network using a various synthesized datasets of images with semi-transparent / opaque visual motifs such as texts, emojis and geometric shapes.

Our network learns to separate the visual motif from the image, by estimating the visual motif matte and reconstructing the latent image. During training, the loss computation

Figure 2: *Watermark removal examples*. Top: input images, middle: reconstruction results, bottom: enlarged patches.

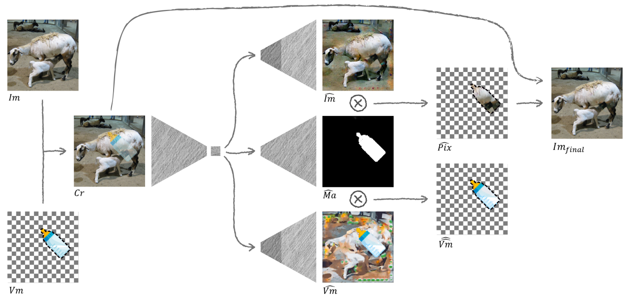

Figure 3: *Method overview.* The network consists of one encoder and three decoders. The top and bottom decoder branches reconstruct the background image and the overlaid visual motif, respectively. The middle branch estimates the mask of the visual motif. The final output is generated by using the mask to select pixels from either the input image or the reconstructed image.

	Translation PSNR / SSIM	Pert. + Opc. PSNR / SSIM	Scale + Rot. PSNR / SSIM
CFM[2]	24.16 / 0.976	N/A	N/A
MMR[1]	37.41 / 0.977	33.07 / 0.966	N/A
SDB[7]	34.64 / 0.973	34.82 / 0.972	34.69 / 0.972
SIRF[8]	31.18 / 0.970	32.87 / 0.970	32.90 / 0.969
Ours	**38.46 / 0.986**	**38.08 / 0.986**	**37.63 / 0.983**

Table 1: Watermark removal comparisons.

uses the input image and the visual motif as ground-truth to train an encoder and decoder networks. Our network encodes the corrupted image into a latent representation, which is decoded by three parallel decoder branches: one for estimating the latent image, motif matte and motif image. The final image is generated by using the estimated motif matte to select pixels from either the input image or the reconstructed image. See Figure 3 for an overview.

3. Experiments

We compare our method to four algorithms [2, 1, 7, 8] in a semi-transparent visual motif removal task, Our test images were embedded with a single unseen visual motif and are divided into three groups with increasing levels of matting deformation. In the first group, the visual motif has a fixed size. For the second group, we added perturbations and varying opacity to the motif blending. For the third group, variations in size and rotation were added. The results (see Table 1) are measured by comparing the reconstructed images to the ground truth one under PSNR and SSIM.

	Light Font Test PSNR / SSIM	Bold Font Test PSNR / SSIM
Sh-CNN[5]	31.879 / 0.9522	28.436 / 0.9118
FoE[6]	38.155 / 0.9886	33.360 / 0.9675
EPLL[9]	38.672 / 0.9884	33.377 / 0.9675
Ours	**39.079 / 0.9890**	**34.676 / 0.9710**

Table 2: Quantitative inpainting results for images corrupted by lighter and bolder font.

We also tested our method in a blind inpainting setting: the network has no explicit priors on the background image or the mask of the corrupted regions. We evaluate our network on two levels of inpainting regions. At each test, 100 images were tiled with random black texts of the font Helvetica light (bold) for the first (second) test. We compare our method to several state-of-the-art inpainting methods [5, 6, 9]. Table 2 summarizes the quantitative results.

References

[1] Tali Dekel, Michael Rubinstein, Ce Liu, and William T Freeman. On the effectiveness of visible watermarks. In *CVPR*, pages 2146–2154, 2017.

[2] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):228–242, 2008.

[3] Jaesik Park, Yu-Wing Tai, and In So Kweon. Identigram/watermark removal using cross-channel correlation. In *CVPR*, pages 446–453. IEEE, 2012.

[4] Soo-Chang Pei and Yi-Chong Zeng. A novel image recovery algorithm for visible watermarked images. *IEEE Transactions on Information Forensics and Security*, 1(4):543–550, 2006.

[5] Jimmy SJ Ren, Li Xu, Qiong Yan, and Wenxiu Sun. Shepard convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909, 2015.

[6] Stefan Roth and Michael J Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205, 2009.

[7] Jie Yang, Dong Gong, Lingqiao Liu, and Qinfeng Shi. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In *ECCV*, September 2018.

[8] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. *CVPR*, 2018.

[9] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *ICCV, 2011*, pages 479–486. IEEE, 2011.