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