

Restoring An Image Taken Through a Window Covered with Dirt or Rain

Cheng Guan

Abstract

Photographs taken through a window are often compromised by dirt or rain present on the window surface. Common cases of this include pictures taken from inside a vehicle, or outdoor security cameras mounted inside a protective enclosure. At capture time, defocus can be used to remove the artifacts, but this relies on achieving a shallow depth-of-field and placement of the camera close to the window. Instead, we present a post-capture image processing solution that can remove localized rain and dirt artifacts from a single image. We collect a dataset of clean/corrupted image pairs which are then used to train a specialized form of convolutional neural network.

1. Introduction

There are many situations in which images or video might be captured through a window. A person may be inside a car, train or building and wish to photograph the scene outside. Indoor situations include exhibits in museums displayed behind protective glass. Such scenarios have become increasingly common with the widespread use of smart phone cameras. Beyond consumer photography, many cameras are mounted outside, *e.g.* on buildings for surveillance or on vehicles to prevent collisions. These cameras are protected from the elements by an enclosure with a transparent window.

Such images are affected by many factors including reflections and attenuation. However, in this paper we address the particular situation where the window is covered with dirt or water drops, resulting from rain. As shown in Fig. 1, these artifacts significantly degrade the quality of the captured image.

The classic approach to removing occluders from an image is to defocus them to the point of invisibility at the time of capture. This requires placing the camera right up against the glass and using a large aperture to produce small depth of-field. However, in practice it can be hard to move the camera sufficiently close, and aperture control may not be available on smart phone cameras or webcams.

In this paper we instead restore the image after capture, treating the dirt or rain as a structured form of image noise. Our method only relies on the artifacts being spatially com-



Figure 1. A photograph taken through a glass pane covered in rain, along with the output of our neural network model, trained to remove this type of corruption. Their regular size and appearance of the rain makes it difficult to remove with existing methods. This figure is best viewed in electronic form.

pact, thus is aided by the rain/dirt being in focus - hence the shots need not be taken close to the window.

Image denoising is a very well studied problem, with current approaches such as BM3D [2] approaching theoretical performance limits [3]. However, the vast majority of this literature is concerned with additive white Gaussian noise, quite different to the image artifacts resulting from dirt or water drops.

1.1. Related Work

Learning-based methods have found widespread use in image denoising, *e.g.* [1, 4]. These approaches remove additive white Gaussian noise (AWGN) by building a generative model of clean image patches. In this paper, however, we focus on more complex structured corruption, and address it using a neural network that directly maps corrupt images to clean ones; this obviates the slow inference procedures used by most generative models.

Neural networks have previously been explored for denoising natural images, mostly in the context of AWGN, *e.g.* Jain and Seung [5], and Zhang and Salari. Algorithmically, the closest work to ours is that of Burger *et al.* [2], which applies a large neural network to a range of non-AWGN denoising tasks, such as salt-and-pepper noise and JPEG quan-

tization artifacts.

2. Approach

To restore an image from a corrupt input, we predict a clean output using a specialized form of convolutional neural network . The same network architecture is used for all forms of corruption; however, a different network is trained for dirt and for rain. This allows the network to tailor its detection capabilities for each task.

2.1. Network Architecture

Given a noisy image x , our goal is to predict a clean image y that is close to the true clean image y^* . We accomplish this using a multilayer convolutional network, $y = F(x)$. The network F is composed of a series of layers F_l , each of which applies a linear convolution to its input, followed by an element-wise sigmoid (implemented using hyperbolic tangent). Concretely, if the number of layers in the network is L , then

$$F_0(x) = x \quad (1)$$

$$F_l(x) = \tanh(W_l * F_{l-1}(x) + b_l), l = 1, 2 \dots L-1 \quad (2)$$

$$F(x) = \frac{1}{m} (W_L * F_{L-1}(x) + b_L), l = 1, 2 \dots L-1 \quad (3)$$

Here, x is the RGB input image, of size $N \times M \times 3$. If n_l is the output dimension at layer l , then W_l applies n_l convolutions with kernels of size $p_l \times p_l \times n_{l-1}$, where p_l is the spatial support. b_l is a vector of size n_l containing the output bias (the same bias is used at each spatial location).

References

- [1] B.A.Olshausen and D.J.Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 1997.
- [2] H. Burger, C. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with BM3D. In *CVPR*, 2012.
- [3] A. Levin and B. Nadler. Natural image denoising: Optimality and inherent bounds. In *CVPR*, 2011.
- [4] S.C.Zhu and D.Mumford. Prior learning and gibbs reaction-diffusion. *IEEE TPAMI*, 1997.
- [5] V.Jain and S.Seun. Natural image denoising with convolutional networks. In *NIPS*, 2008.