

RUNet: A Robust UNet Architecture for Image Super-Resolution

Xiaodan Hu¹ Mohamed A. Naiel¹ Alexander Wong¹ Mark Lamm² Paul Fieguth¹

¹Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada

²Christie Digital Systems Canada Inc., Kitchener, ON, Canada

{x226hu, mohamed.naiel, a28wong, pfi eguth}@uwaterloo.ca, mark.lamm@christiedigital.com

Abstract

Single image super-resolution (SISR) is a challenging ill-posed problem which aims to restore or infer a high-resolution image from a low-resolution one. Powerful deep learning-based techniques have achieved state-of-the-art performance in SISR; however, they can underperform when handling images with non-stationary degradations, such as for the application of projector resolution enhancement. In this paper, a new UNet architecture that is able to learn the relationship between a set of degraded low-resolution images and their corresponding original high-resolution images is proposed. We propose employing a degradation model on training images in a non-stationary way, allowing the construction of a robust UNet (RUNet) for image super-resolution (SR). Experimental results show that the proposed RUNet improves the visual quality of the obtained super-resolution images while maintaining a low reconstruction error.

1. Introduction

Modern state-of-the-art single image super-resolution (SISR) methods have been deep learning-based methods [1–5], which have demonstrated significant reconstruction quality improvements. For example, generative adversarial network-based SR methods [1, 2] have been able to generate realistic results, but these methods suffer from unstable training. On the other hand, convolutional neural network (CNN) based methods [3–5] have shown effectiveness in learning a nonlinear relationship between low and high resolution images. However, such methods [3–5] underperform when handling images with non-stationary degradations. One of the reasons is that a majority of these methods [3, 4] leverage a Bicubic down-sampling image degradation model for approximating the true degradation [6], which is not true in many practical scenarios such as projector resolution enhancement. Furthermore, such network

architectures [3–5] are limited in their ability to learn complex non-stationary degradations.

Motivated by this, we propose a robust UNet (RUNet) architecture for image super-resolution to learn how to treat different image contents in a way that achieves better super-resolution results. More specifically, the proposed RUNet leverages long-range connections to improve learning capabilities, and leverages a degradation model based on spatially varying degradations that force the network to learn handling spatially non-stationary image degradations. Experimental results show that the proposed RUNet offers super-resolution images with improved visual quality while maintains a low reconstruction error.

2. Proposed Method

The proposed resolution enhancement scheme consists of a degradation module and a new UNet architecture as shown in Figure 1. During training, a set of input training images of high resolution are first downsampled by a factor of two in both directions and then blurred, at random, using a Gaussian filter. Next, every blurred image is upsampled by a factor of two in both x and y directions using the Bicubic interpolation for initializing the proposed network. For training the proposed network, every upsampled blurred image and its corresponding image at the original resolution are used. In testing, given a low-resolution input image, an upsampling operator by a factor of two is performed in both the x and y directions, and then the trained network is used to predict the enhanced high-resolution image.

2.1. Network Architecture

The proposed RUNet architecture consists of a number of convolutional layers, batch norms, ReLU activation functions, and tensor operations as shown in Figure 1. Unlike the conventional UNet [7] architecture, the left path shown in Figure 1 consists of a sequence of blocks each followed by a tensor addition operation to feed forward the same block input to the subsequent block, so-called residual block [4]. This allows the network to learn more complex structures. In order to efficiently upscale the low-resolution image, the sub-pixel convolutional layers [8] are

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC-CRD), Christie Digital Systems Inc., and the Ontario Centres of Excellence (OCE-VIPII).

