Very Low Resolution Image Classification

Adam Vest[1] apvest01@louisville.edu
Muhammadabdullah Jamal[2] a_jamal@knights.ucf.edu
Boqing Gong[2] bgong@crcv.ucf.edu

[1] University of Louisville [2] University of Central Florida

Abstract

The problem of classifying low resolution images is both highly challenging as low resolution images carry very limited information and incredibly important for developing robust, resolution-agnostic classification systems. In the past researchers have attempted to solve the low resolution image classification problem directly, that is, by training classifiers on a dataset of low-res images (e.g., CIFAR-10), or by employing a simple image resizing algorithm such as bicubic interpolation. In this work, we take a different approach and explore the use of a super-resolution network to transform input LR images into the high resolution image domain before classification. We show that using the above technique, we are able to improve low resolution image classification accuracy.

1. Introduction

The problem of classifying low resolution images remains highly challenging despite the widespread improvements of image classification networks. Many low resolution image datasets are relatively simple (e.g., CIFAR-10 or MNIST); containing very few classes, large inter-class variation, and little intra-class variation. While classifier performance on such simple datasets is usually quite good, classification accuracy on more complex low resolution datasets is significantly worse (e.g., the state of the art model in [7] suffers a nearly 5x performance degradation on CIFAR-100 as compared to CIFAR-10). However, when classifiers are trained on more complex (i.e., more classes and reduced inter-class variation) high resolution datasets such as ImageNet, they still achieve comparable accuracies to classifiers trained on simpler low resolution datasets. What causes high resolution classifiers to perform better on complex datasets than their low resolution counterparts? Does the increased information carried by high resolution images aid in the classification task or are performance benefits due simply to training set size? If the classification task is aided by increased image resolution, how can we use this knowledge to improve low resolution image classification performance? In this work, we explore these questions in detail and develop a simple framework that allows classification systems to exploit the increased information carried by higher resolution images to improve low resolution image classification performance.

1.1. Related Work

The problem of low resolution image classification was one of the first image classification problems studied by the computer vision community. The literature on classification of such low resolution image datasets as MNIST and CIFAR-10 is vast but generally relies on the same basic methodology: In order to perform classification of low resolution images, classifiers are trained or fine-tuned directly on low resolution images. While we share the same goal of low resolution image classification, we take a different approach to previous literature in that we first apply a super-resolution algorithm to map the input low resolution images into a higher dimensional space before classification. A basic flow diagram of our proposed low resolution image classification pipeline can be seen in Figure 1.

Closer conceptually to our work is that of Hennings-Yeomans et. al [3], Zou et. al [8], and others who have used super-resolution as a utility to improve facial recognition performance on very low resolution images. Importantly, [3] demonstrates that learning a super-resolution and facial recognition model concurrently allows for increased recognition performance on low resolution images. Despite these basic similarities, there are some fundamental differences between our approaches including (1) our use of deep super-resolution and classifier networks instead of simple Tikhonov super-resolution [1] and a linear classifier, (2) the proposed joint training schemes, and (3) our goal of enabling both high and low resolution image classification in a single unified system as opposed to learning a model only capable of classifying low resolution images.

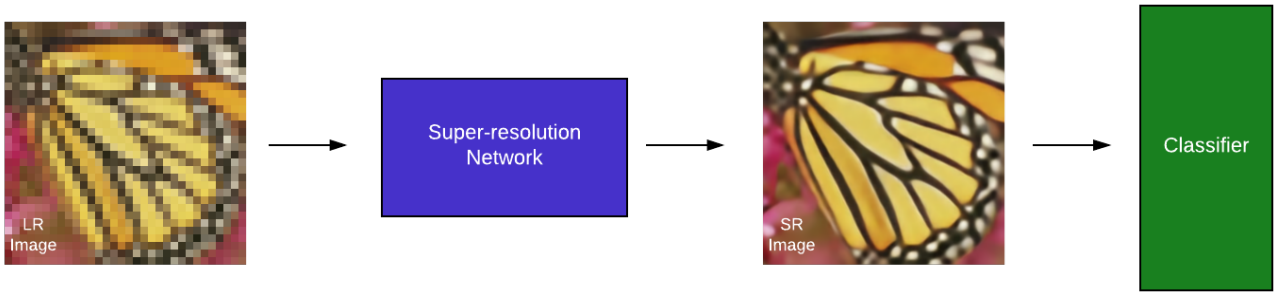


Figure 1: Proposed low resolution image classification pipeline.

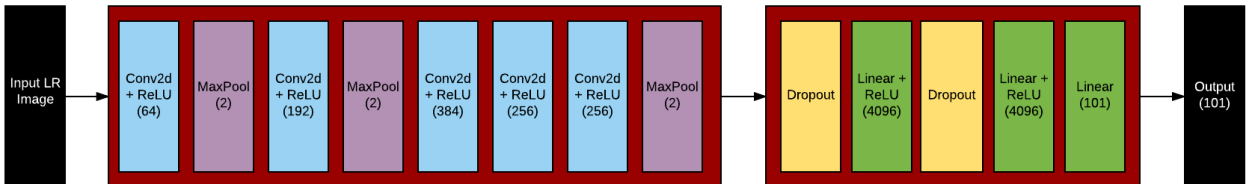


Figure 2: Low resolution image classifier architecture.

2. Experiments

2.1. Setup

All experiments were run on the Caltech101 dataset which was divided into 3 splits of 50%, 25%, and 25% for training, validation and testing while maintaining relative class importance. All images were preprocessed by scaling to 256, taking a 224x224 center crop, and (for low resolution images) downsampling to 28x28 via bicubic interpolation (scaling factor of 8x).

In order to test our hypothesis that classifiers can exploit the increased information in higher resolution images, we utilize two image resizing algorithms to map input low resolution images into the high resolution image domain before classification: bicubic interpolation and the state-of-the-art deep super-resolution model SRResNet [5]. As reported in Ledig et. al, SRResNet significantly outperforms bicubic interpolation in terms of image reconstruction quality as measured by PSNR, SSIM, and other metrics, allowing us to examine the impact of image reconstruction quality on classifier performance. Our implementation of SRResNet is the same as the basic model used in Ledig et. al: the network has 16 residual blocks and was trained to upscale images by a factor of 4 using a mean squared error content loss. For more information on SRResNet training, see [5].

For our experiments, we employ 2 basic types of classifiers: (1) models trained on low resolution images and (2) models trained on high resolution images. For fair comparision, we use AlexNet as the base model for all classifiers. Given the striding and pooling used in the AlexNet architecture, the low resolution classifiers were modified to use a standard 3x3 kernel and a zero-padding and stride of 1 to prevent the max pooling operation from decreasing the image size to zero; however, the number and type of layers, the number of feature maps in each convolutional layer, and the number of neurons in each fully connected layer are the same as in the high resolution classifiers. A simple diagram of the low resolution classifier architecture can be seen in Figure 2. All models were initialized with pretrained weights from ImageNet (low resolution classifiers were initialized with weights from a downsampled version of ImageNet) and fine-tuned to the Caltech101 dataset. All models were trained using an initial learning rate of 10^{-3} which was decayed by a factor of 10 if validation accuracy failed to improve by .5% over three epochs. Adhering to standard practice, fine-tuned layers were given a decreased learning rate of $\frac{1}{20}$th the base learning rate. Training ended for all models when validation accuracy failed to improve by .001% over the course of 10 epochs.

The evaluation metrics chosen for our models are the Top-1 and Top-5 classification accuracies. The low resolution models were evaluated on low resolution inputs only whereas the high resolution models were evaluated on both ground truth high resolution images and images generated after the application of a super resolution algorithm to low resolution images. In the context of reported results and subsequent discussion, low resolution models evaluated on low resolution images will be referred to as LR classifiers, high resolution models evaluated on high resolution images will be referred to as HR classifiers, and high resolution

Model	Top-1 Accuracy	Top-5 Accuracy
LR Classifier	50.07%	70.80%
HR Classifier	85.68%	96.57%
BC SR Classifier	67.60%	88.24%
SRResNet SR Classifier	78.63%	92.68%

Table 1: Top-1 and Top-5 classification accuracies of LR, HR, and SR classifiers.

models evaluated on low resolution images super-resolved by a super-resolution algorithm will be referred to as SR classifiers.

2.2. Results and Discussion

The preliminary results for 4x upscaling are reported in Table 1. All reported results are the average accuracies of three models. These early results support our hypotheses that we can improve low resolution image classification performance using super-resolution and that classification performance improves with the quality of image reconstruction.

As expected, the HR classifiers perform the best of the three model types (HR, LR, SR) acheiving a Top-1 accuracy of nearly 86%. It is likely that these HR models benefit from both high-quality, high resolution training and testing images as well as the use of models pretrained on the large, high resolution ImageNet dataset as a base initialization. One surprising result was the extremely poor performance by the low resolution classifier which achieved a Top-1 accuracy of just 50%. We conjecture that the severe performance degradation of these classifiers is due to the reduced information in low resolution images and having to train the classifier from scratch instead of initializing with pretrained weights from ImageNet. Immediate increases were seen in the classification accuracy of low resolution images when we first applied a super-resolution algorithm to map them into a higher dimensional space and then classified them using the HR classifier. This method immediately improved classification accuracy by 17% and 28% for low resolution images super-resolved by bicubic interpolation and SRResNet respectively. The large increase in classification accuracy on super-resolved images is likely due to both the quality of the HR classifier and the increased amount of discriminative information contained in the super-resolved images. Given the dramatic increase in classification performance between low resolution and super-resolved images, the preliminary results support our first hypothesis that low resolution image classification can be improved through the use of super-resolution.

Another interesting result is the large gap between the performance of the classifier on images super-resolved by different algorithms. Specifically, using SRResNet, which achieved state-of-the-art super-resolution performance as measured by image quality metrics such as PSNR and SSIM, to super-resolve images instead of bicubic interpolation yields an 11% performance boost, suggesting that the quality of image reconstruction is an important determining factor for super-resolved image classification performance.

3. Future Work

Most of our future work on this topic will be on the problem of learning super-resolution and classification networks that enable the use of a single classifier to perform classification of both low and high resolution images. Stated formally, given a high resolution image classifier in the desired domain, f_c, we would like to find a new classifier $f_{\tilde{c}} = f_c + u$ for some vector of small offset weights u that maximizes classification performance on low resolution images while also maintaining high resolution image classification accuracy. Cast in this manner, the problem of designing a multiple-resolution image classifier is a modified domain adaptation problem where we desire to maximize performance on a given task in both the target *and* source domains. There are a few options we may pursue to attempt to solve this problem including:

1. **Fine-tuning**. Fine-tuning SRResNet using classification loss on super-resolved images should force SRResNet to produce super-resolved images that are easier to classify. Another approach would be to fine-tune the high resolution classifier on images super-resolved by SRResNet. The number of update iterations would likely have to be small in order to avoid decreasing high resolution classification performance significantly.

2. **Residual loss**. The residual loss function has been used to improve the performance of facial recognition models trained on a large source dataset when fine-tuned on a much smaller target dataset. We may need to adapt the loss function to ensure high resolution classification performance does not degrade though it is possible that the regularization term could be suffi-

cient to prevent significant decreases in performance in the source domain.

3. **Adversarial examples inspired techniques**. Training with adversarial examples has been shown to force classifiers to learn a better representation of a class in that they become resistant to small changes in the input image that are imperceptible to the human eye. It is possible that adversarial example training alone could create classifiers that are more resistant to the small changes in input images created by imperfect super-resolution algorithms. Some methods that could be worth exploring are defensive distillation [6], fast gradient sign [2], and learning with a strong adversary [4]. Additionally, we may be able to modify these techniques to produce perturbations that cause the input high resolution training images to become more similar to their corresponding low-resolution image counterpart. Specifically, we should be able to generate perturbations using a modified form of the fast gradient sign method in [2]:

$$\eta = -\epsilon sign(\nabla_{I^{HR}} MSE(I^{HR}, I^{SR})) \qquad (1)$$

where η is the perturbation, ϵ is a small constant, and $\nabla_{I^{HR}} MSE(I^{HR}, I^{SR})$ refers to the gradient of the mean squared error loss function with respect to the input high resolution training image. These super-resolution perturbations could also be generated using any method capable of pushing the level set boundaries of our high resolution classifier f_c towards those of some super-resolved image classifier f_{SR}, so more existing andnew techniques of generating these perturbations could also be explored.

References

[1] M. Elad and A. Feuer. Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images. *IEEE Transactions on Image Processing*, 1997.

[2] I. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. *ICLR*, 2015.

[3] P. Hennings-Yeoman, S. Baker, and B. Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. *CVPR*, 2008.

[4] R. Huang, B. Xu, D. Schuurmans, and C. Szepesvari. Learning with a strong adversary. *ICLR*, 2016.

[5] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *ArXiv e-prints*, 2016.

[6] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami. Distillation as a defense to adversarial perturbations against deep neural networks. *IEEE Symposium on Security and Privacy*, 2016.

[7] S. Xie, R. Girshick, P. Dollr, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. *CVPR*, 2017.

[8] W. Zou and P. Yuen. Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, 2012.