Super Resolution to Improve Classification Accuracy of Small Images

Christian Martinez-Nieves
Stanford University
chris151@stanford.edu

Luke Jaffe
Stanford University
jaffe5@stanford.edu

Shiv Sundram
Stanford University
shiv1@stanford.edu

Introduction

- We apply single-image super resolution (SISR) as a pre-processing step for solving the image classification problem on image datasets.
- We expect that using super-resolution as a pre-processing step will help yield higher classification accuracy.
- Super-resolution methods will often use either per-pixel loss functions, or perceptual loss functions [1].
- Previous approaches to the SISR task have focused on improving image quality as measured by human perception or pixel signal-to-noise ratio (PSNR).
- Our goal is to improve machine perception of imagery by enhancing discriminative features.

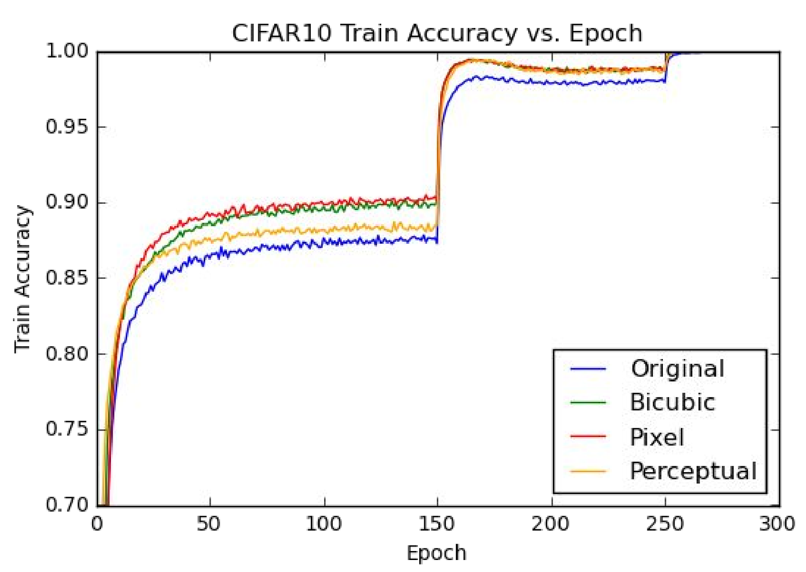

Figure 1: Comparison of training accuracy for CIFAR-10 images using bicubic, pixel loss and perceptual loss based upsampling.

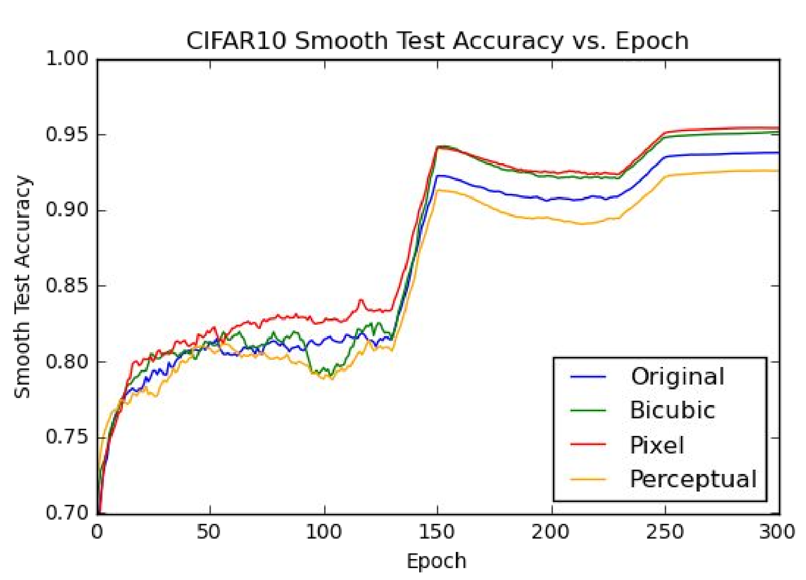

Figure 2: Comparison of test accuracy for CIFAR-10 images using bicubic, pixel loss and perceptual loss based upsampling.

Problem Statement

- Determine if super-resolution can be used as a pre-processing step to improve image classification accuracy.
- Carry out experiments using general image classification datasets, like CIFAR-10, as well as domain specific datasets, like Food-101.
- Evaluation will be based on comparing classification accuracy between the original image datasets` and the super-resolved image datasets.

Datasets

- CIFAR-10: 60,000 32x32 images with 10 basic classes, and 6000 images per class.
- STL-10: 100,000 unlabeled 96x96 images, similar to CIFAR-10 domain.
- IMAGENET: 15 million labeled high-resolution images in over 22,000 categories.
- Microsoft COCO: 300,000 labeled high-resolution images with 80 categories.
- Food-101: 101,000 images with 101 food categories, each class having 1,000 images.

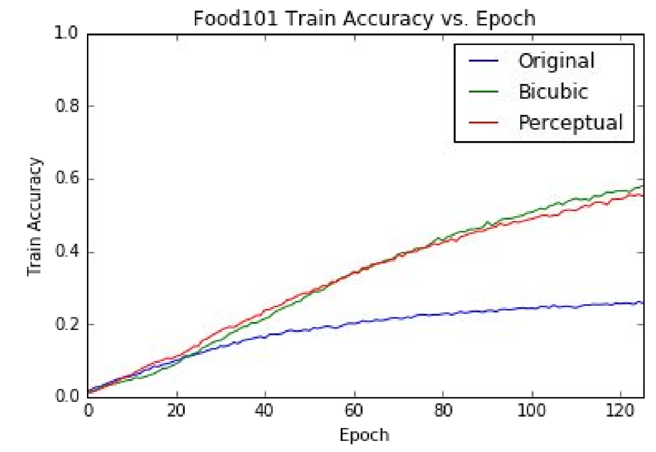

Figure 3: Comparison of training accuracy for Foods-101 images using bicubic, and perceptual loss based upsampling.

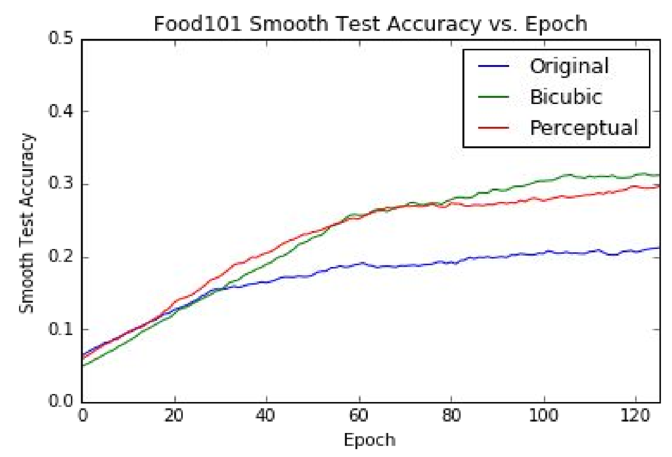

Figure 4: Comparison of test accuracy for Foods-101 images using bicubic, and perceptual loss based upsampling.

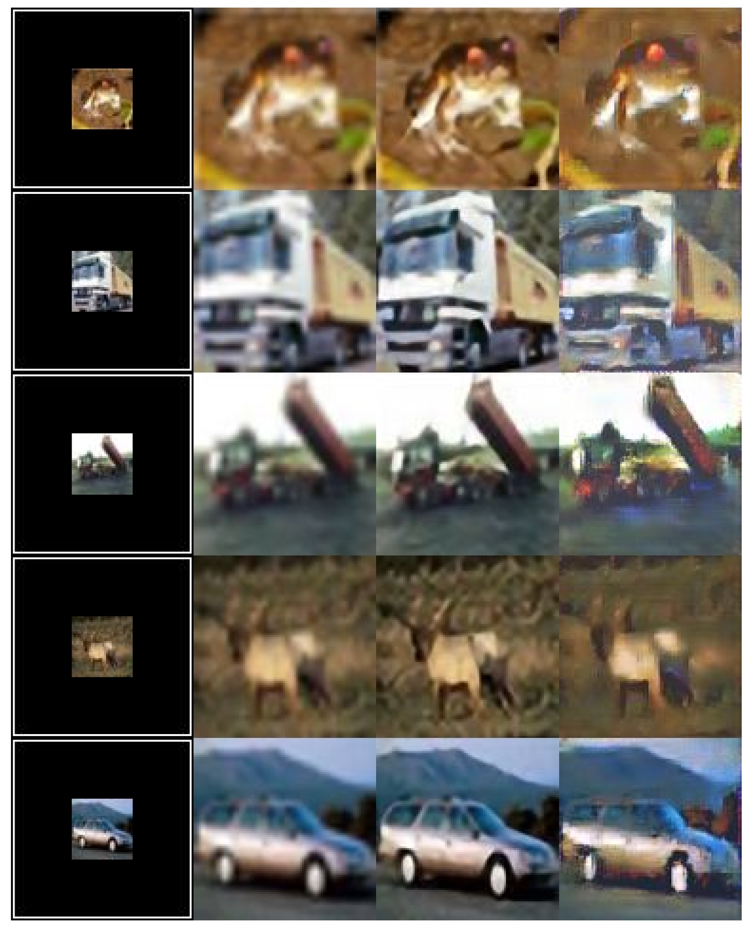

Figure 5: Original CIFAR10 images are juxtaposed against bicubic upsampled (2nd col), pixel based upsampled (3rd col), and perceptual based upsampled (4th col).

Figure 6: Original Food 101 images are juxtaposed against bicubic upsampled (2nd col), and perceptual based upsampled (3rd col).

Approach

- We investigated three approaches for upsampling.
  - Shi et al. [2] , uses a standard per-pixel loss function to train an array of upscaling filters. Trained on STL10
  - Johnson et al. [3], which uses perceptual loss functions (also used for fast style transfer) and histogram matching. Trained on COCO. Loss network is a pre-trained VGG network trained on Imagenet
  - bicubic interpolation, mainly to be used as a control.
- In all cases, we used a VGG network to perform classification (with some extra layers added to handle the larger, enhanced inputs).

Results

- Super-resolution adds features that make an image more aesthetically pleasing to a human, but it does not necessarily add features that make classes more separable to a classifier
- Final classification accuracy did not improve significantly with super-resolution used as a preprocessing step.
- However, Johnson's super-resolution method resulted in significantly higher test accuracy at the beginning of training
- Classification accuracy for bicubic and super-resolved images is higher than original images'; bicubic and super-resolved images are fed into VGG with more layers (to handle larger inputs

Conclusion

- Since classification accuracy did not improve significantly when using super-resolution, our results did not support our hypothesis
- However, the original hypothesis may still hold when working with domain specific datasets.
- Since super resolution did result in higher test accuracy at the beginning of training a CNN, it may still be possible that using super resolution can help speedup training under certain circumstances.

References

[1] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." *European Conference on Computer Vision*. Springer International Publishing, 2016.

[2] Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

[4] Dong, Chao, et al. "Learning a deep convolutional network for image super-resolution." *European Conference on Computer Vision*. Springer International Publishing, 2014.

[5] Glasner, Daniel, Shai Bagon, and Michal Irani. "Super-resolution from a single image." *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009.