ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 4, No. 2, XX-YY (2019)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Object Classifications by Image Super-Resolution Preprocessing for Convolutional Neural Networks

Bokyoon Na[1], Geoffrey C Fox[2]

[1]*Dept. of Computer Engineering, Korea Polytechnic University, 429-793, Korea*

[2]*School of Informatics, Computing, and Engineering, Indiana University, 47408, USA*

ARTICLE INFO

Article history:
Received:
Accepted:
Online:

Keywords:
Object classification
Super-resolution
Convolutional neural networks
Machine learning

ABSTRACT

Blurred small objects produced by cropping, warping, or intrinsically so, are difficult to detect and classify. Therefore, much recent research is focused on feature extraction built on Faster R-CNN and follow-up systems. In particular, RPN, SPP, FPN, SSD, and DSSD are the layered feature extraction methods for multiple object detections and small objects. However, super-resolution methods, as explored here, can improve these image analyses working on before or after convolutional neural networks. Our methods are focused on building better image qualities into the original image components so that these feature extraction methods become more effective when applied later. Our super-resolution preprocessing resulted in better deep learning in the number of classified objects, especially for small objects when tested on the VOC2007, MSO, and COCO2017 datasets.

1. Introduction

This paper is an extension of work originally presented in IEEE international conference on Big Data 2018 [1]

Since Krizhevsky et al. [2] introduced custom designed CNN (Convolutional Neural Network) architectures, there have been many methods to increase the rate of object classification and detection for CNNs. [3], [4], [5][6], [7][8] have shown performance to be increased compared to the shallow learning in neural networks. Nowadays, with many blur-less or just slightly blurred images, CNNs classify objects with around 90 percent classification rates. Moreover, the algorithms in [9], [10] introduced much faster detection times, an increased number of classes, and object segmentation in addition to the softmax and linear regression algorithms.

Recently, promising research to reduce false and missing detections with CNN networks includes generative adversarial networks (GAN) [11][12], [13], GAN networks with reinforcement learning [14], and Capsule networks [15]. GAN networks address the issues from adversarial noise. Sara Sabour et al. [15] have got considerably better results than CNNs on MultiMNIST with smaller sized training data sets. However, Capsule networks, consisting of a group of neurons learning to detect a particular object within a region of the image [15], do not perform as well as CNNs on larger images. This occurs even though Capsule networks address the deficiencies of CNNs using pose information (such as precise object position, rotation, size, and so on). Also, note that Capsule networks require far more computing resources than CNNs.

Even though this research has improved CNNs in many ways, there are still detection failures caused by blurred or low-resolution images. For example, 100 images from [16] were randomly chosen and preprocessed at three times lower resolution than the original images to build cropped or warped images, which are called the regions of interest (ROI). Then they were interpolated by bilinear or bicubic algorithm, and finally, they were tested by [4]. These interpolation methods may cause aliasing effects on the image and make the identified region of interest larger. Therefore, new methods are needed to build less aliased and high-resolution ROIs from original images in CNNs.

Since they were introduced, the most popular image size for CNN has been around 256x256 pixels. Examples of the Spatial Pyramid Pooling network, Faster R-CNN, and ConvNet have been tested for the significance of the image size. Also, in [17][17], [19], [20], it is shown that a recurrent neural network model is capable of extracting information from an image by selecting a sequence of regions and processing every selected region at high resolution. We see that object detection from ROIs is a popular topic in CNN research.

Alex Krizhevsky et al. [2] described the number of categories and designed their network with 1,000 object classes from ImageNet. [3], [6] classified 21 object classes from PASCAL VOC 2007 [16]. There is a 200 class dataset in ILSVRC2013/ ILSVRC2017, which contains several datasets, and especially, the ImageNet dataset contains 1000 classes, and Omniglot contains 1623 classes. But much research is often done with VOC2007 20+1 classes even though there are datasets with more than several hundred classes. [9] trained and tested with the 80 object classes in the fastest object detection speed. As future research, more than 200,000 object categories may need to be considered to distinguish objects such as people.

In practical applications, there are many low-quality image processing systems such as surveillance camera systems, car black-box systems, or even mobile phone cameras for taking pictures of long-distance. For example, in surveillance systems, it may not easy to increase the capacities for better image qualities because of their storage capacities, dark conditions on-camera image sensors, and night visioning. Especially, current night visioning algorithms do not have the procedures for good image qualities. Also, in-car black box systems, moving vibrations, and the requirements of low electric power consumption may cause bad image compressions or lack of fast zooming or focusing devices. Finally, mobile phone pictures are blurred or have small target objects according to the limit of the software zooming algorithms.

Given this background, we will introduce our research to improve object classification rates with improvements through super-resolution algorithms for convolution neural networks.

2. Related Research

Compared to the previous CNN algorithms, [3], [21] have shown new algorithms of new parameters related to lower layers of the neural networks. J.R.R. Uijlingsvan de Sande et al. [17] introduced a new feature extraction algorithm for object recognition. The paper [3] also improved the CNN "region proposals" technique. Their new methods speed up object detections allowing the image dataset to run in almost real-time. Using region proposals, the Fast YOLO model in [9] processed 155 frames per second.

Keiming He et al. [6] had the best results for training and testing with the maximum image size of 392 pixels as a shorter side of the input image because they had variable size image datasets from VOC 2007 and ImageNet. Also, they showed results indicating that scale matters in the classification processing. Thus, they suggested that the spatial pyramid pooling model supports different image sizes in convolution layers, that work with various image sizes while the standard fully connected layer requires a fixed image size. With the Caltech101 image dataset, they found that detections of objects had better performances among several scaled datasets. They noticed that this is mainly because the detected objects usually occupy large regions of the whole images. They evaluated cropped or warped images and got lower accuracy rates than the same model on the undistorted full images.

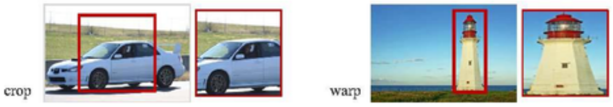


Figure 1: Cropping and warping. This image is from [6]

From the above discussion of related work, we identified two motivations to get better performances in object recognitions, which are "region of interests" and "the high resolution for a region of interests which is a cropped area from the input image." It means that high-resolution image cropping is our approach, and super-resolution methods achieve this.

In [22], super-resolution is used to construct high-resolution images from several observed low-resolution images or from a single low-resolution image. This increases the high-frequency components and removes the degradations caused by image processing at low resolution.

Let X denote the desired high-resolution image and Y_k be the kth low-resolution observation. Assume the imaging system captures K low-resolution frames of X, where the low-resolution data are related to the high-resolution scene X by

$$Y_k = D_k H_k F_k X + V_k, k = 1, 2, \dots, k,$$

where F_k encodes the motion information for the kth frame, H_k models the blurring effects, D_k is the down-sampling operator, and V_k is the noise term. These linear equations can be rearranged into a linear system

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix} = \begin{bmatrix} D_1 H_1 F_1 \\ D_2 H_2 F_2 \\ \vdots \\ D_k H_k F_k \end{bmatrix} \mathbf{X} + \underline{\mathrm{V}} \qquad (1)$$

or equivalently, $\underline{Y} = \mathrm{M}\mathbf{X} + \underline{V}$.

In real imaging systems, these matrices are unknown and need to be estimated from the available low-resolution observations.

Unlike the above image observation model, statistical approaches in [23] relate the super-resolution reconstruction steps stochastically to the optimal reconstruction. The high-resolution image and motions among low-resolution inputs are regarded as stochastic variables. Therefore, the super-resolution reconstruction is cast into a fully Bayesian framework using a sum rule and the Bayes theorem:

$$\mathrm{X} = \arg\overset{max}{\underset{x}{}} \Pr(X|\underline{Y}) \qquad (2)$$

$$= \arg\overset{max}{\underset{x}{}} \int_{v,h} \Pr(\underline{Y}|X, M(v,h) \Pr(X) \Pr\big(M(v,h)\big)\, dv.$$

More details on the procedures may be found in [23]. Here $\Pr(\underline{Y}|X, M(v,h)$ is the likelihood of the data, Pr(X) is the prior term on the desired high-resolution image and $\Pr\big(M(v,h)\big)$ is a prior term for motion estimation. Finally, if $M(v,h)$ is given or estimated as M, then X can be obtained as the Maximum a Posteriori (MAP) estimate, or even X can be reduced to the simplest maximum likelihood (ML) estimator, which can be treated by the Expectation-Maximization (EM) algorithm.

In [24], two types of super-resolution algorithms are introduced: single image-based and multiple images based. In the multiple-image super-resolution algorithm, two low-resolution images of the same scene are fed in as input, and one adds a registration algorithm to find the transformation between them. The single image super-resolution algorithm employs a training step to learn the relationship between a set of high-resolution images and their low-resolution counterparts, and this relationship is used to predict the missing high-resolution details of input images. There are three steps, which are the registration (motion estimation), the restoration, and the interpolation to construct a high-resolution image from a low-resolution image or multiple low-resolution images [25].

In addition to the EM algorithm for single image super-resolution, [25] present methods using a mixture of experts (MoE), which are anchor-based local learning approach, sparse coding, and deep convolution neural networks, respectively. Christian Ledig et al. [29] proposed a GAN method using generator networks and discriminator networks to recover photorealistic natural images with minimization of the mean squared reconstruction error.

In this paper, we will apply a single image super-resolution method to get better detection rates from single image dataset.

3. Image Resolution Improvements

To improve the input image qualities, there are two kinds of approaches. The first is to apply the super resolution directly into the input images, then CNN takes the super resolution image as the input images. The second is to apply the super resolution only for bounding boxes. The second method may reduce the needed computational effort.

Shaoqing Ren et al. [4] implemented the interpolation algorithm for two purposes. The first is for fixing the input image size from variable input images, and the second one is for a selective search to extract feature maps from input images. The detector extends networks for windowed detection by a list of crops or selective search (or EdgeBoxes) feature maps and then interpolates them to fix the size into one of the bounding boxes. Thus, this method may have more bounding boxes for small objects as shown in Figure 2 and described in detail in section 4 of [6]. We leave for future research, the use of an extended image size directly adopted by the super-resolution method. Instead, we explore in this paper, a super-resolution method to increase the image size to around 492x324 pixels.

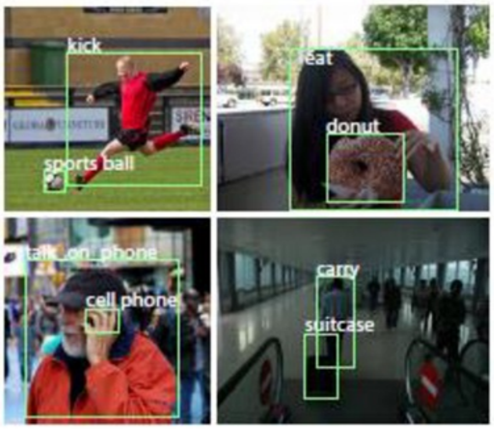


Figure 2: Images with either full-size objects or small-sized objects. These examples are from [30]

Thus, we will describe super-resolution methods first to improve the image quality in pre-processing before the input layer of CNN; then, we describe how the preprocessed image can be classified by a CNN.

3.1. Super-resolution as Pre-processing

In image analysis, there have been distinct improvements related to convolutional neural networks. However, we will focus on the pre-processing of image samples, which is by super-resolution methods. For example, if a cropped image with 166x110 pixels is extracted from an image with 640x480 pixels, the cropped image is usually too small to feed into the input layer of CNN models to get good results in object recognition. Thus, we will add a super-resolution method before the input layer of CNN or for bounding boxes.

We propose a Mixture of Experts (MoE) model to solve problems given in the anchor-based local learning, which are optimizing the partition of feature space and reducing the number of anchor points. The MoE model will be solved by the Expectation-Maximization (EM) algorithm. The objective of an MoE model is to partition a large complex set of data into smaller subsets via a gating function.

Being an expert means that each of the model component classifiers or regressors is highly trained, component regressors, W_i, and as an anchor point, v_j have relations such as:

$$\min_{\{v_1,v_2,\ldots,v_k;\, w_1,w_2,\ldots,W_k\}} \Sigma_{j=1}^{N} \sum_{i=1}^{K} c_{ji} ||h_j - W_i l_j||^2, \qquad (3)$$

where l_j means a low-resolution path, h_j is the corresponding high-resolution patch, v_j is the nearest anchor point for l_j and c_{ji} is a continuous scalar value which represents the degree of membership of l_j.

However, to improve the computational efficiency and the competitive image quality of the anchor-based local learning method of multiple regressors, a mixture of experts, which is one of the conditional combined mixture models [31] [32], is proposed here.

In the mixture of experts model training, a maximum likelihood estimation should be solved iteratively by the EM algorithm. Every iteration, the posterior probabilities are calculated for patches, and then we get the expectation of the log-likelihood as an E-step. During M-step, anchor points and regressors are updated, which is a softmax regression problem. After training, super-resolution images can be constructed by collecting all the patches from regressed low-resolution patches and averaging the overlapped pixels.

To compare the performance of the super-resolution method with interpolation methods, bilinear and bicubic interpolated images will be built in addition to the images produced by the super-resolution method.

3.2. Object classification

As the object classification model, we implemented a convolution neural network based on the Faster R-CNN [4] approach because it supports variable image sizes as the input data of CNN and sliding widow proposal scanning for the convolution network. Above all, the detection speed is fast enough and almost real-time. As mentioned earlier, Faster R-CNN uses region proposals to detect an object, and the ratio of the size of the region proposal to the whole image is critical in detecting the object. As shown in Figure 2, sometimes objects from small bounding boxes are ignored because the object may be deformed or aliased at cropping or warping. With a better quality of image, CNN will make a higher quality of cropped or warped bounding boxes. This approach allows us to distinguish our proposed method compared to other ways of interpolating the data.

The CNN is implemented based on Intel® Xeon CPU 2.30GHz and 4 NVIDIA Tesla K80 GPU boards. Each GPU board has a memory of 12GB and two GPUs.

Our CNN has several convolutional layers to support a region proposal network (RPN) in addition to the conventional CNN layers [2][8]. Also, these convolution layers are shared with object detection networks in the same way as [6]. Thus, our model generates a deep CNN, which has many convolutional and pooling layers.

As multiple-scale prediction schemes for the training of Faster R-CNN, there are methods based on multiple-images, on multiple-filters, and on multiple-anchors that are given in [4]. With multi-image schemes, images are resized at multiple scales and feature maps are computed for each scale. Even though this scheme is time-consuming, our super-resolution method can resize images to get better prediction scores. However, to lower the use of computational resources, for fast detections, and for feature sharing in fully convolution layer, we choose the the multi-anchors scheme.

Additionally, we address the memory usages of CAFFE by constraining the number of images read from the image dataset. In our deep CNN model allowing multiple scaled image sizes, the number of region proposals is w (width of the image) x h (height of the image) x k (the number of anchors of a region proposal). It means that memory space for the only region proposal network is simply required severely. Therefore, we constrain the number of images in hidden layers to be less than or equal to 20 simultaneous image. We also looked at a version of the Keras model which is using TensorFlow and GPUs. but new model had similar issues as CAFFE and we do not present results based on it.

For model training, we use pre-trained model parameters taken from Faster R-CNN implementation. Faster R-CNN had training, validating and testing with PASCAL VOC 2007 of 9000 images. Also, we tested with Microsoft's COCO dataset. As a result, we consider that the parameters of our model are reliably determined.

For each proposal from an image, an object is roughly considered, and the proposal size is selected as one of several predefined window sizes. Thus, this patch of an image may be interpolated. If instead of interpolation methods the super-resolution method is adopted, any variable sizes of input images with variable sizes or shapes can be supported with better cropped or warped regions even though the Faster R-CNN supports images with variable sizes.

We will not use the super-resolution method to improve the patch image qualities here. We work with low-resolution images from the input dataset, processing super-resolution on these images, then detecting objects. We will consider this extension in future research.

4. Performance Evaluations

There are two kinds of popular file formats for the input datasets to detect objects from images with CNN networks. Therefore, we considered several image file formats to get better image quality for pre-processing, training, and testing images. While most of the image datasets are built with images based on the JPG file format, this did not seem to offer as good results in super-resolution processing compared with the BMP format. As a result of this observation, we decided to use BMP format and to convert images from a file format of JPG into BMP format as a pre-processing procedure. Also, we scale down image size by 3 times smaller in each width and height to build images with lower quality, instead of directly collecting images by cropping or warping images. Then, the image is taken to be further processed with bilinear, and bicubic interpolations, and the super-resolution method.

In our model, we implement the super-resolution procedure based on the approach given in [26]. Rather than training with less than 50 images as in most of the published super-resolution models, we have trained our model based on their initial parameters and with 100 PASCAL VOC2007 images. There is no overfitting with this number of images for training. With random images from PASCAL VOC2007 [16][33] and test images from [34], and additionally with COCO dataset images, we have tested our super-resolution model. We could not find any differences between these testing methods. Therefore, we will test the object classification with 520 images randomly extracted from VOC2007 and 1224 images from COCOs MSO [35].

In Figure 3, we illustrate that the dataset has a different number of objects compared to our intuition. For example, the first image of Figure 3 is labeled with no object, but our proposed model detected an object or objects, such as shown in the figure. The second image of Figure 3 shows that a sofa is detected even though the image is labeled with no object. Thus, we decide not to use the given label from the dataset. The number of persons in the third image of Figure 3 is detected as larger than the labeled count. It means that we need to count the number of objects manually for every image in the dataset to get the precision and recall.

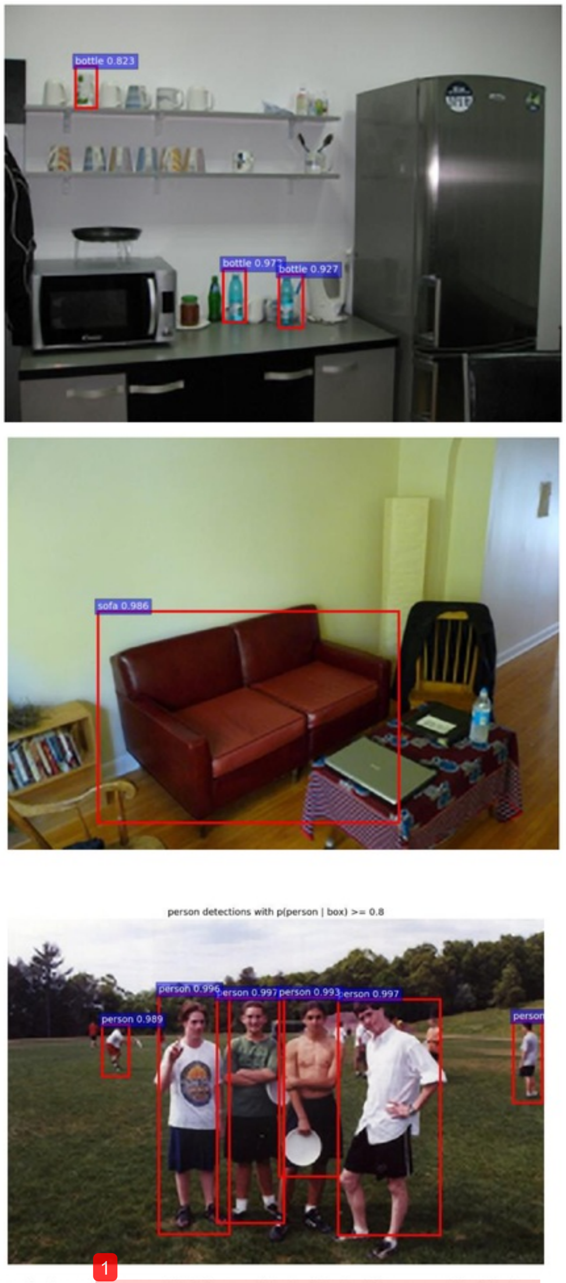


Figure 3: Some Images which have objects detected by our method but are labeled with no objects or a lesser number than we find.

4.1. Comparison of Output Pictures

Many images from the PASCAL VOC 2007 have multiple objects, as shown in the results given in 0with 3 different pre-processing models, which are a bilinear interpolation, bicubic interpolation, and our super-resolution method. Our model is set with a learning rate of 0.001 and detection scores with 0.8 or higher.

The first column (green color bars) shows the number of detected objects in each class, as given by bilinear interpolation. For the second column (blue color bars), we show for bicubic interpolation, the number of objects detected, and the detection rate. The third column (red color bars) shows results for our super-resolution approach.

As shown in Table 1, our model detects more objects than the other two models, but the average detection scores are not very different. It means that our model makes improved input images, which are fed into the CNN, which is able to increase the number of detections. Therefore, objects with near or over threshold scores (80% or above) are added, even though they did not get the scores over threshold value in bilinear or bicubic interpolation models. Figure 4 **Error! Reference source not found.**shows the cumulative number of detected objects and comparison between them.

With the dataset of Microsoft MSO as another image data for testing of object classification, we present results in Figure 5 and Table 2. As with the VOC2007 dataset, our proposed model detects more objects than the two other models. Again high scored objects are detected in all three models: bilinear, bicubic, and our model, but almost all of the low score objects are detected only in our super-resolution model. As with VOC2007, this tends to depress the scores of the super-resolution model.

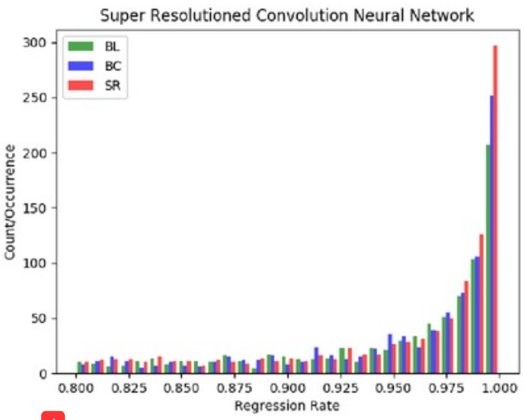


Figure 4: Histogram for convolution network models with 3 different image pre-processing methods on randomly extracted images from VOC 2007 dataset

Table 1: Detected objects on 3 different pre-processing and convolution neural networks with the PASCAL VOC2007 dataset. #tp is the number of true positives correctly predicted. The column labeled mean is the average probability from the method.

	Bilinear		Bicubic		Our Method	
Classes	#tp	mean	#tp	mean	#tp	mean
aeroplane	25	0.9569	29	0.9472	27	0.9787
bicycle	16	0.9602	17	0.9697	18	0.9520
bird	18	0.9204	25	0.9178	30	0.9500
boat	16	0.9330	18	0.9260	20	0.9276
bottle	19	0.9222	17	0.9208	15	0.9331
bus	26	0.9626	25	0.9639	26	0.9588
car	93	0.9662	100	0.9671	104	0.9710
cat	11	0.9427	10	0.9665	11	0.9739
chair	25	0.9273	34	0.9368	45	0.9320
cow	11	0.9149	12	0.9216	16	0.9385
dining table	7	0.9317	7	0.9420	12	0.9193

dog	37	0.9559	36	0.9657	35	0.9565
horse	30	0.9560	34	0.9574	37	0.9694
motorbike	13	0.9467	13	0.9630	16	0.9581
person	405	0.9546	432	0.9575	466	0.9590
potted plant	10	0.9467	12	0.9194	18	0.8767
sheep	15	0.9275	13	0.9380	14	0.9227
sofa	7	0.9191	8	0.9175	8	0.9475
train	7	0.9193	4	0.9276	4	0.9572
tvmonitor	25	0.9653	25	0.9729	26	0.9479
Total	816	0.9415	871	0.9450	**948**	**0.9465**
misclassified	25		21		**39**	

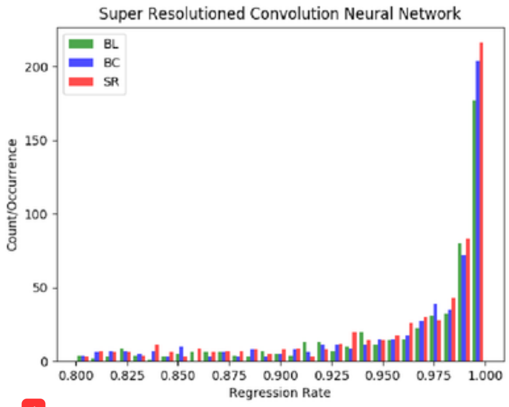


Figure 5: Histogram for convolution network models with 3 different image pre-processing methods on randomly extracted images from MSO dataset

Table 2: Detected objects on 3 different pre-processing and convolution neural networks with the MSO dataset. #tp is the number of true positives correctly predicted. The column labeled mean is the average probability from the method.

	Bilinear		Bicubic		Our Method	
Classes	#tp	mean	#tp	mean	#tp	mean
aeroplane	5	0.9650	6	0.9300	9	0.9263
bicycle	1	0.9992	2	0.9112	1	0.9978
bird	50	0.9512	59	0.9521	75	0.9627
boat	1	0.8301	1	0.9771	1	0.9713
bottle	23	0.9074	24	0.9062	21	0.9117
bus	3	0.9954	3	0.9954	3	0.9871
car	16	0.9641	19	0.9558	18	0.9499
cat	11	0.9639	10	0.9829	11	0.9540
chair	19	0.9299	20	0.9270	22	0.9393
cow	5	0.9452	6	0.9488	7	0.9584
dining table	3	0.9410	4	0.8927	4	0.9072
dog	42	0.9559	47	0.9636	50	0.9478
horse	11	0.9728	11	0.9726	15	0.9362
motorbike	4	0.9487	4	0.9529	4	0.9588
person	302	0.9715	318	0.9718	340	0.9719
potted plant	4	0.9023	5	0.8756	9	0.9035
sheep	1	0.8780	1	0.9353	4	0.9064
sofa	3	0.9204	3	0.9099	6	0.9261
train	6	0.9746	7	0.9529	8	0.9522
tvmonitor	6	0.9720	6	0.9804	10	0.9248
Total	516	0.9444	556	0.9447	**618**	**0.9447**
misclassified	86		82		**92**	

The main change in COCO 2017 test dataset [36] is that instead of an 83K/41K training/validation split, based on community feedback, the split is now 118K/5K for training/validation. The same exact images are used, and no new annotations for detection/key-points are provided. However, new in 2017 is annotations on 40K training images (a subset of the full 118K training images from 2017) and 5K validation images. Also, for testing, in 2017, the test set only has two splits (dev / challenge) instead of the four splits (dev / standard / reserve / challenge) used in previous years. With this dataset, we have the results shown in Figure 6 and Table 3. The total number of detected objects of ours is much larger than for the bilinear or bi-cubic models. However, the mean probabilities of detected objects are similar or a little bit lower. It means that many objects are not detected with bilinear or bi-cubic interpolation methods but these are detected with our super-resolution method. These newly detected objects in our method which cause the average probability to decrease, presumably have probabilities, which are lower than 0.80 in the bilinear or bi-cubic methods.

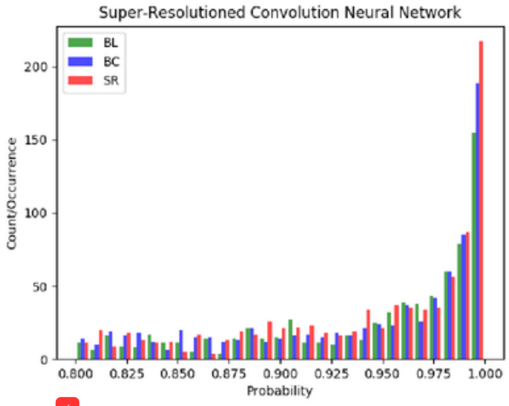


Figure 6: Histogram for convolution network models with 3 different image pre-processing methods on randomly extracted images from COCO 2017 dataset

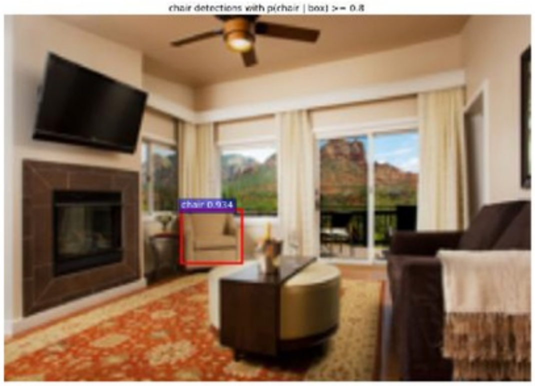
Table 3: Detected objects on 3 different pre-processing and convolution neural networks with COCO 2017 test dataset. #tp is the number of true positives correctly predicted. The column labeled mean is the average probability from the method.

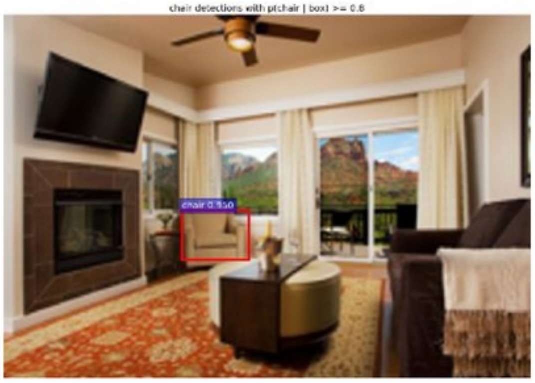
	Bilinear		Bicubic		Our Method	
Classes	#tp	mean	#tp	mean	#tp	mean
aeroplane	9	0.9652	10	0.9502	11	0.9684
bicycle	1	0.8821	2	0.9012	3	0.9250
bird	5	0.9471	10	0.9066	17	0.9277
boat	5	0.9002	5	0.8908	7	0.8896
bottle	20	0.9116	25	0.9100	19	0.9160
bus	15	0.9725	16	0.9590	15	0.9456
car	24	0.9503	29	0.9251	28	0.9388
cat	6	0.9274	7	0.9235	10	0.9075
chair	26	0.9193	28	0.9169	36	0.9174
cow	12	0.9190	14	0.9287	14	0.9384
dining table	10	0.8989	9	0.9049	9	0.9063
dog	12	0.9184	12	0.9183	12	0.9047
horse	7	0.9219	8	0.9293	6	0.9402

motorbike	11	0.9604	14	0.9550	18	0.9368
person	411	0.9547	460	0.9521	498	0.9534
potted plant	16	0.8806	11	0.8824	15	0.9002
sheep	2	0.9585	2	0.9880	2	0.9977
sofa	3	0.9398	3	0.9015	2	0.8760
train	9	0.9358	10	0.9269	9	0.9427
tvmonitor	26	0.9461	24	0.9457	25	0.9544
Total	734	0.9305	805	0.9258	**869**	**0.9293**
misclassified	104		106		**113**	

4.2. Big vs. Small ROI Pictures and Their Detection Rates

As mentioned in the previous section, if objects are big enough compared to the size of the image which is containing the object proposal, objects from interpolated images with bilinear or bicubic methods are identified satisfactorily and sometimes may have better performance than our proposed model. Better performance means the object may have a higher probability score that the object is the specific class. However, our proposed model has much better results with objects from small bounding boxes or small ratio of object size to the size of the image, which contains the object. In Figure 7, our proposed model detects a chair, which is a quite small object in the image, while the other two models did not detect this 'chair' object. Even though the other chair is detected in all three models, our proposed model has a somewhat higher score. Thus, we can conclude our model can help to improve image qualities in image analyses. Especially with video surveillance camera systems, our model may help to detect more objects. Therefore, we are interested in this research with surveillance camera systems as a future topic.





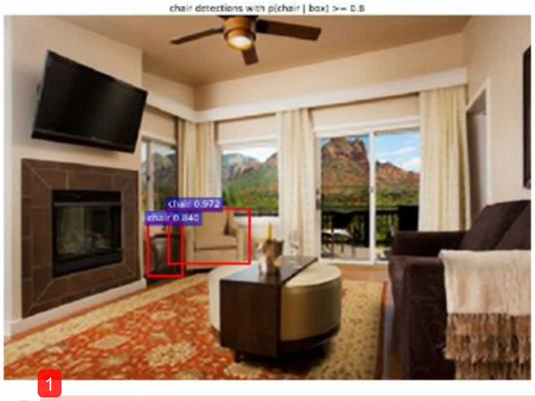


Figure 7: Comparison of small object detection through Bilinear, Bicubic, and super-resolution models

5. Conclusion

Our proposed model appears particularly powerful in three scenarios; firstly, where there are relatively small objects in large pictures; secondly, where there is warping in the region proposals approach; and finally, with object detection from cropped images. Commonly, there are many noisy video images from surveillance camera systems, especially with night vision systems. Our method will remove a significant amount of aliased or mosaicked areas in these images, and so help to detect more objects. We compared our scheme with other approaches on three datasets showing an increased object in each case.

In future work, we will implement the super-resolution method confined to the bounding box areas. This will reduce the needed computational resources and allow us to use this method in real-time processing [37] and achieve better object recognition in this application. We will demonstrate this capability using modern streaming software environments [38].

6. Acknowledgments

This work was supported by the 2017 sabbatical year research grant of the Korea Polytechnic University. Geoffrey Fox was partially supported by NSF CIF21 DIBBS 1443054 and NSF CINES 1835598. We thank the Futuresystems group for the computational resources.

References

[1] Bokyoon Na, Geoffrey Fox, "Object Detection by a Super-Resolution Method and a Convolutional Neural Networks", IEEE International Conference on Big Data (Big Data) (2018)

[2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. "ImageNet Classification with Deep Convolution Neural Networks". In Advances in Neural Information Processing Systems (NIPS), (1097-1105). (2012).

[3] Girshick Ross. "Fast R-CNN". arXiv: 1504.08083v2 [cs.CV] 27 Sep 2015.

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks". arXiv:1506.01497v3[cs.CV] 6 Jan 2016.

[5] David Berthelot Schumm, Luke Metz Thomas. "BEGAN: Boundary Equilibrium Generative Adversarial Networks". arXiv: 1703.10717v2 [cs.LG]. (2017).

[6] Keiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recogniton". arXiv:1406.4729v4. (2015).

[7] Ross Girshick, Jeff Donahua, Trevor Darrell, Jitendra Malik, "Region-Based Convolution Networks for Accurate Object Detection and Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 38, NO.1*, 2016.

[8] Karen Simonyan, Zisserman Andrew. "Very Deep Convolutional Networks for Large-Scale Image Recognition". ICLR. (2015).

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-time Object Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788

[10] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick, "Mask R-CNN", The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961-2969

[11] Ghosh Nachum and Debiprasad Ofir. https://www.quora.com. "What-are-Generative-Adversarial-Networks-GANs".

[12] Tim Salimans Goodfellow, Wojciech Zaremba, Vicki Cheung Ian. "Improved Techniques for Training GANs". arXiv:1606.03498v1. (2016).

[13] Takeru Miyato M Dai, Ian Goodfellow Andrew. "Adversarial Training Methods for Semi-Supervised Text Classification". ICLR. (2017).

[14] Ian J. Goodfellow Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courvill, Yoshua BengioJean. "Generative Adversarial Nets". arXiv:1406.2661v1. (2014).

[15] Sara Sabour, Nicholas Frosst, Geoffrey E. Hinton, "Dynamic Routing Between Capsules", Advances in Neural Information Processing Systems 30 (NIPS 2017)

[16] M. Everingham Van Gool, C. K. I. Williams, J. Winn, and A. ZissermanL. "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results". (2007).

[17] Achanta R., Estrada F., Wils P., Süsstrunk S. "Salient Region Detection and Segmentation". In: Gasteratos A., Vincze M., Tsotsos J.K. (eds) Computer Vision Systems. ICVS 2008. Lecture Notes in Computer Science, vol 5008. Springer, Berlin, Heidelberg. (2008)

[18] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. "Recurrent Models of Visual Attention". arXiv:1406.6247v1[cs.LG] 24 Jun 2014.

[19] Jimmy Lei Ba, Volodymyr Mnih, and Koray Kavukcuoglu. "Multiple Object Recognition with Visual Attention". arXiv:1412.7755v2[cs.LG] 23 Apr 2015.

[20] Karol Gregor Danihelka, Alex Graves, Danilo Jimenez Rezende, Daan Wierstra Ivo. "DRAW: A Recurrent Neural Network For Image Generation". arXiv:1502.04623v2[cs.CV] 20 May 2015.

[21] J.R.R. Uijlingsvan de Sande, T. Gevers, and A.W.M. Smeulders K.E.A. "Selective Search for Object Recognition". IJCV. (2012).

[22] Michael Elad, Arie Feuer, "Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images", IEEE Transactions on Image Processing, Vol. 6, No. 12, December 1997

[23] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma, "Image Super-Resolution Via Sparse Representation", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO. 11, pp2861, NOVEMBER 2010

[24] Nasrollahi Kamal, Guerrero Escalera Sergio, Rasti Pejman, Anbarjafari Gholamerza, Baro Xavier, J. Escalante Hugo, Moeslund B. Thomas. "Deep Learning based Super-Resolution for Improved Action Recognition". In International Conference on Image Processing Theory, Tools and Applications (IPTA) IEEE Signal Processing Society. (2015).

[25] Brian C. Tom, Aggelos K. Katsaggelos, Nikolas P. Galatsanos. "Reconstruction of a high resolution image from registration and restoration of low resolution images". In Proceedings of IEEE International Conference on Image Processing, pages 553-557, 1994

[26] Kai Zhang Wang, Wangmeng Zuo, Hongzhi Zhang, Lei ZhangBaoquan. "Joint Learning of Multiple Regressors for Single Image Super Resolution". IEEE Singnal Processing Letters. Vol. 23 No. 1. (2016).

[27] Zhaowen Wang Liu, Jianchao Yang, Wei Han, Thomas Huang Ding. "Deep Networks for Image Super Resolution with Sparse Prior". ICCV. (2015).

[28] Chao Dong Change Loy, Kaiming He, Xiaoou Tang Chen. "Image Super Resolution Using Deep Convolution Networks". arXiv:1501.00092v3.v (2015).

[29] Christian Ledig Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe ShiLucas. "Photo-Realistic Single Image Super Resolution Using a Generative Adversarial Network". arXiv:1609.04802v5[cs.CV]. (2017).

[30] Georgia Gkioxari, Ross Girshick, Piotr Dollár, Kaiming He, "Detecting and Recognizing Human-Object Interactions", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8359-8367

[31] Bishop Christopher. "Pattern Recognition and Machine Learning". Springer. (2006).

[32] Zhang Zhang, Baoquan Wang, Wangmeng Zuo, Hongzhi ZhangKai. "Joint Learning of Multiple Regressors for Single Image Super-Resolution". IEEE Signal processing letters, Vol. 23, No.1. (2016).

[33] Olga Russakovsky Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei Jia. "ImageNet Large Scale Visual Recognition Challenge". arXiv:1409.0575. (2014).

[34] Olga Russakovsky Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei Jia, "ImageNet Large Scale Visual Recognition Challenge 2017". (2017).

[35] Zhang Ma, Shuga Sameki, Mehrnoosh Sclaroff, Stan Betke, Margrit Lin, Zhe Shen, Xiaohui Price, Brian Much, Radom Jianming. "Salient Object Subitizing". IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015).

[36] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár. "Microsoft COCO: Common Objects in Context", arXiv:1405.0312 [cs.CV]

[37] Xinyuan Huang, Geoffrey C. Fox, Sergey Serebryakov, Ankur Mohan, Pawel Morkisz, Debojyoti Dutta, "Benchmarking Deep Learning for Time Series: Challenges and Directions", Stream-ML workshop at IEEE Big Data Conference, Los Angeles CA December 10, 2019

[38] Vibhatha Abeykoon, Supun Kamburugamuve, Kannan Govindrarajan, Pulasthi Wickramasinghe, Chathura Widanage, Niranda Perera, Ahmet Uyar, Gurhan Gunduz, Selahattin Akkas, Gregor Von Laszewski, and Geoffrey Fox, "Streaming Machine Learning Algorithms with Big Data Systems", Stream-ML workshop at IEEE Big Data Conference, Los Angeles CA December 10, 2019

19M-10-375 V2

ORIGINALITY REPORT

62%	56%	56%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dsc.soic.indiana.edu Internet Source	50%
2	Submitted to Universidad Politecnica Salesiana del Ecuado Student Paper	2%
3	www.ifp.illinois.edu Internet Source	2%
4	Submitted to Rochester Institute of Technology Student Paper	1%
5	Bokyoon Na, Geoffrey C Fox. "Object Detection by a Super-Resolution Method and a Convolutional Neural Networks", 2018 IEEE International Conference on Big Data (Big Data), 2018 Publication	1%
6	Kai Zhang, Baoquan Wang, Wangmeng Zuo, Hongzhi Zhang, Lei Zhang. "Joint Learning of Multiple Regressors for Single Image Super-Resolution", IEEE Signal Processing Letters, 2016	1%

	Publication	
7	Submitted to International Islamic University Malaysia Student Paper	1%
8	www-levich.engr.ccny.cuny.edu Internet Source	<1%
9	www.ijert.org Internet Source	<1%
10	"Super-Resolution Imaging", Springer Nature, 2002 Publication	<1%
11	Submitted to ABV-Indian Institute of Information Technology and Management Gwalior Student Paper	<1%
12	S. Lertrattanapanich, N.K. Bose. "High resolution image formation from low resolution frames using delaunay triangulation", IEEE Transactions on Image Processing, 2002 Publication	<1%
13	A. Zisserman. "Bayesian Methods for Image Super-Resolution", The Computer Journal, 10/27/2007 Publication	<1%
14	Pulasthi Wickramasinghe, Supun Kamburugamuve, Kannan Govindarajan, Vibhatha Abeykoon et al. "Twister2: TSet High-	<1%

	Performance Iterative Dataflow", 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), 2019 Publication	
15	Submitted to Modern Education Society's College of Engineering, Pune Student Paper	<1%
16	Submitted to University of Durham Student Paper	<1%
17	elib.uni-stuttgart.de Internet Source	<1%
18	"Image and Graphics", Springer Science and Business Media LLC, 2019 Publication	<1%
19	persagen.com Internet Source	<1%
20	Submitted to University College London Student Paper	<1%
21	Yi Ma. "Image super-resolution as sparse representation of raw image patches", 2008 IEEE Conference on Computer Vision and Pattern Recognition, 06/2008 Publication	<1%
22	F.P. Ferrie. "Comparison of Super-Resolution Algorithms Using Image Quality Measures", The	<1%

	3rd Canadian Conference on Computer and Robot Vision (CRV 06), 2006 Publication	
23	Submitted to University of Southampton Student Paper	<1%
24	Entregado a Chulalongkorn University el 2011-10-07 Student Paper	<1%

Exclude quotes	On	Exclude matches	< 3 words
Exclude bibliography	On		

19M-10-375 V2

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8