CS230

Super-Resolution of Low-Quality Dashcam Images for Realtime Pothole Detection

Moo Jin Kim †
Department of Computer Science
Stanford University
moojink@stanford.edu

Sharan Ramjee †
Department of Computer Science
Stanford University
sramjee@stanford.edu

Abstract

In this paper, we present a novel approach to enable the use of low-resolution cameras for automated pothole detection through the use of super-resolution (SR) via Super-Resolution Generative Adversarial Networks (SRGANs). We then proceed to establish the baseline pothole detection performances on low-resolution and high-resolution dashcam images using a YOLO network. Finally, we demonstrate and analyze the pothole detection performance gain achieved over the baselines by applying super-resolution on the low-resolution dashcam images.

1 Introduction

Publicly available pothole detection datasets (1; 2) largely consist of high-resolution images used for training object detection models (3). However, given the expensive nature of high-resolution dashcams, dashcam owners instead opt for cheaper, low-resolution dashcams (especially for bicycles and scooters). As such, object detection models do not perform well in practice due to the domain-mismatch between the high-resolution training data and low-resolution test data. In order to address this issue, we propose a Super-Resolution Generative Adversarial Network (SRGAN) (4) based deep learning pipeline that enhances the resolution of dashcam footage in order to enable accurate pothole detection on low-resolution dashcams. The inputs to the pipeline are low-resolution dashcam images, and the outputs are the predicted bounding boxes, along with confidence scores, of any potholes present in the upscaled version of the images.

2 Related Works

Past approaches for automated pothole detection involved equipment such as accelerometers and gyroscopes (5), wireless IoT sensors (6; 7), and thermal imaging cameras (8). Due to the cost and complexity of setting such devices up, (9) leveraged deep learning for end-to-end pothole detection given dashcam footage, testing models like SSD and YOLOv3 and finding that YOLOv3 achieved the quickest and most reliable pothole detection. (3) also achieved fast and accurate pothole detection with various YOLOv3 architectures. However, no past work has evaluated the effects of super-resolution specifically on dashcam images of potholes. Kim et al. used an SRGAN to enhance low-resolution images of vehicles caught by CCTV cameras before they were fed into a CNN-based vehicle model classifier (10), finding that upscaling low-resolution images from 224x224px to 896x896px led to a significant increase in classification accuracy when compared to the baseline performance without the SRGAN. These results indicate that super-resolution improves classification of large objects such as vehicles, but our motivation is to determine how an SRGAN might improve performance not in classification but in object detection, where the objects to be detected are also much smaller: potholes.

† Equal contribution

3 Dataset

We used the pothole detection dataset compiled by the Electrical and Electronic Engineering Department at Stellenbosch University (1; 2), which consists of 3888 RGB images of size 3680x2760. Sample images from the dataset are shown in Figure 1. We employed a train-validation-test split of 68%-16%-16% on the 3888 total images. In order to evaluate the performance of our pipeline, we made two copies of the dataset: one downscaled to 720p (1280x720px), which is the low-resolution version of the dataset, and another rescaled to 4K (3840x2160px), which is the high-resolution version of the dataset[1].

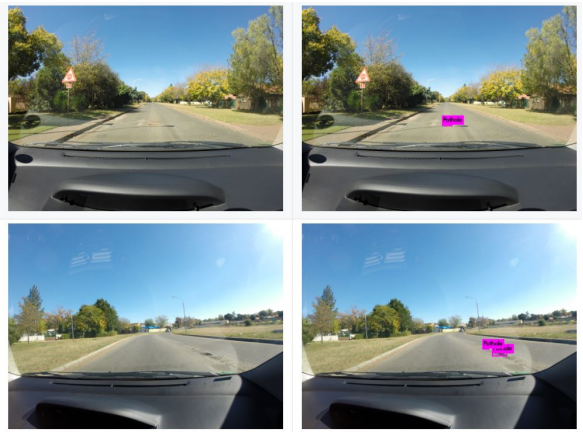

Figure 1: Original dashcam images (left) and their corresponding YOLO model outputs (right)

4 Methods

The pothole detection pipeline consists of an SRGAN (4) followed by a YOLOv4 model (11) where the SRGAN performs super-resolution on the incoming stream of low-resolution dashcam frames to convert them to high-resolution frames, which are then passed on to the YOLO model which performs object detection on the frames to output frames with bounding boxes around the potholes as illustrated in Figure 2.

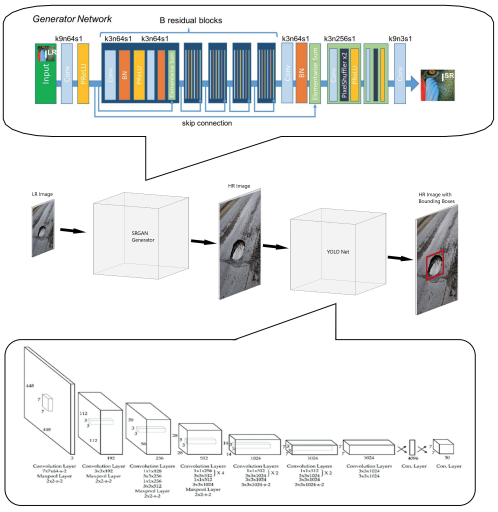

Figure 2: The SRGAN - YOLO Pothole Detection pipeline

[1]As explained later in Section 5, we additionally made a third copy of the dataset that was downscaled to 360p (640x360px) to further demonstrate the impact of super-resolution on even lower image resolutions.

4.1 SRGAN

The SRGAN (4) consists of a generator, which upscales the low-resolution (LR) images to super-resolution (SR) images, and a discriminator, which distinguishes between the high-resolution (HR) and SR images and backpropagates the GAN loss to train the discriminator and generator as observed in Figure 3.

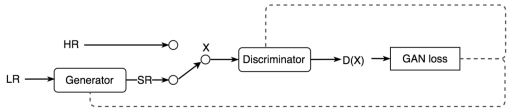

Figure 3: Backpropagation in an SRGAN

The loss function of the SRGAN generator is a summation of the content/reconstruction loss and the adversarial loss:

$$l^{SR} = l^{SR}_{X} + 10^{-3} l^{SR}_{Gen}$$

where the content loss is computed pixel-wise using the Mean Square Error (MSE) between the HR and SR images. The SRGAN uses the perceptual loss that measures the MSE of features extracted by a VGG19 network (12) that has been pre-trained on the ImageNet dataset (13). To match the features, given a specific layer of the VGG19 model, we measure the MSE as:

$$l^{SR}_{VGG/i,j} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

and the adversarial loss as:

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta G}(I^{LR}))$$

The loss function of the SRGAN discriminator is given by the discriminator loss that is ubiquitously used in most GANs (14):

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_D}(I^{LR})))]$$

After employing transfer learning and fine-tuning the SRGAN on the pothole detection dataset, Figure 4 shows that 4x super-resolution to 4K resolution is almost indistinguishable from the original 4K image, and an image super-resolved to 720p also differs appreciably from the 360p image.

Figure 4: Downscaled 360p image (top left), 720p image upscaled from 360p using SRGAN (top right), 4K image upscaled from 720p using SRGAN (bottom left), original 4K image (bottom right)

4.2 YOLO

The YOLO network (15) is a CNN that splits the input image into a grid of cells, where each cell is responsible for predicting whether or not the center of a pothole lies inside of it, and outputs a

vector containing the pothole detection confidence scores of all cells along with their corresponding bounding box predictions. Due to its low latency and high mean average precision (mAP) compared to other YOLO models, we decided to use the YOLOv4 (11) model after adapting the original network architecture while tuning hyperparameters (as described in the next section). We employed transfer learning on a YOLOv4 model that was trained to classify 80 different objects in the 2017 MS COCO dataset (16), initializing the model with these pre-trained weights and then fine-tuning it to detect potholes (a single class) using the aforementioned pothole detection dataset (1; 2). Concretely, we fine-tuned two models until convergence (i.e., until the mAP score started to increase on the validation set): one baseline model trained on low-resolution 720p (1280x720) pothole images, and another model trained on higher-resolution 4K (3840x2160) images[2]. The code with the SRGAN and YOLOv4 implementations used for the experiments are available on GitHub[3].

5 Experiments, Results, and Discussion

For the SRGAN, a variety of hyperparameters were explored, including other GAN architectures such as the RDN (17). Using grid search, the final hyperparameter values that were chosen were: generator, feature extractor, and discriminator loss weights of 0.0, 0.833, and 0.01, respectively. A learning rate with an initial value of 0.0004 was employed with a decay factor of 0.5 and a decay frequency of 30. Finally, we used mean average error, mean squared error, and binary cross-entropy for the generator, feature extractor, and discriminator, respectively, in accordance with (4). For the YOLOv4 model, we used an initial learning rate of 0.001 with a decay factor of 0.1 that decays the learning rate after 36000 mini-batches and again after 40500 mini-batches so that we converged at the local optima without overshooting. We chose a mini-batch size of 64 since this led to faster learning compared to other powers of 2. Finally, we used a momentum of 0.949, in accordance with (11).

To show that our SRGAN-upscaling of dashcam images from 720p to 4K improves pothole detection performance, we first established a baseline model performance by training the YOLO model on a training set consisting of 720p dashcam images and then testing it on a test set also consisting of 720p images. We then trained another YOLO model on the higher-resolution 4K images, and then tested it with three different test sets of different image resolutions: 720p images (to highlight the performance reduction caused by domain mismatch), 4K images upscaled from 720p by the SRGAN (to highlight the performance improvement achieved through the SRGAN), and the original 4K images (to set an upper benchmark on performance). Furthermore, we repeated these experiments with lower resolution images, now using 360p and 720p images and using the SRGAN to upscale from 360p to 720p. The results for both super-resolution experiments are shown below in Table 1.

Table 1: Pothole detection performance evaluation for different input image resolutions

			Evaluation Metrics	
Model	Training set	Test set	**mAP (%)**	F1-score
YOLOv4 (baseline)	720p	720p	64.01	0.67
YOLOv4	4K	720p	60.67	0.64
YOLOv4	4K	**4K (SRGAN)**	**65.84**	**0.68**
YOLOv4 (upper benchmark)	4K	4K (original)	68.79	0.70
YOLOv4 (baseline)	360p	360p	28.49	0.39
YOLOv4	720p	360p	31.09	0.42
YOLOv4	720p	**720p (SRGAN)**	**60.06**	**0.65**
YOLOv4 (upper benchmark)	720p	720p (original)	64.01	0.67

As shown above, the YOLO model achieves a higher mean average precision (mAP) and F1 score with the SRGAN-upscaled 4K images than with the 720p images, surpassing the baseline model that is trained and tested on 720p images. In addition, the performance obtained with the SRGAN images is close to the upper benchmark obtained with the original 4K images.

[2] As explained in the next section, we additionally trained on very low-resolution 360p (640x360 px) images to further demonstrate the impact of super-resolution on pothole detection performance.

[3] https://github.com/sharanramjee/pothole-srgan-yolo

The results with the 360p and 720p settings are similar to those with the 720p and 4K settings, but now the impact of the SRGAN-upscaling is much more pronounced: we achieve significantly higher mAP and F1 scores with the SRGAN-to-YOLO pipeline than with the baseline model trained and tested on 360p images.

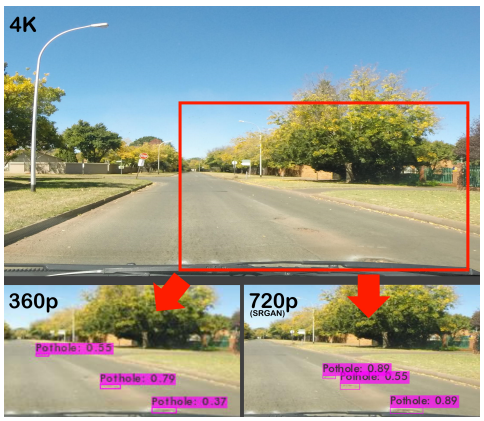

Figure 5: Pothole detection on a 360p image (bottom left) versus on a 720p image (bottom right) upscaled from 360p using the SRGAN. Original 4K image (top) provided for reference.

Figure 5 above illustrates the pothole detection output from the YOLO model given a sample 360p input image and a 720p input image that was upscaled from 360p by the SRGAN. As expected, the YOLO model detects potholes better with the upscaled image; given the 360p image, the baseline model fails to detect one of three potholes and incorrectly detects the shadow of the head of the streetlight as a pothole as the low-resolution makes it difficult to distinguish between these objects.

6 Conclusion and Future Work

Our results clearly indicate an improvement in pothole detection performance via SRGAN-upscaling of low-resolution input images. Very low-resolution 360p test images can be fed into the SRGAN-to-YOLO pipeline to obtain mAP scores over 60 percent, which is surprising considering that potholes are harder to detect than other, more conspicuous objects, such as cars and humans. These findings have various implications: If compute power is more readily available than memory, the driver may record dashcam videos with low-resolutions to save memory since large resolutions lead to diminishing returns, especially if the objects to be detected are larger than potholes. Alternatively, if certain object detection tasks require very fine details for reasonable performance (such as detecting extremely distant objects), then SRGANs can be applied to feed much higher-resolution images to the object detection model to improve performance. In the future, we plan to assess different models, such as SRResNet, FaSTGAN, and PP-YOLO and their effects on pothole detection performance while attempting to improve the inference latency of the pothole detection pipeline.

7 Contributions

7.1 Moo Jin Kim

Fine-tuned three different versions of the YOLOv4 model (trained on the 360p, trained on 720p, and trained on 4K training sets as described previously). Evaluated the YOLO models' performances and obtained the metrics shown in the "Experiments, Results, and Discussion" section.

7.2 Sharan Ramjee

Implemented and fine-tuned the SRGAN for upscaling the low-resolution images in the pothole detection dataset. Wrote utils for running inference on videos and for downscaling the high-resolution 4K images to various low-resolution image sizes (360p, 720p, and 1080p). Created and documented the GitHub for maintaining the `pothole-srgan-yolo` codebase.

References

[1] S. Nienaber, M. J. Booysen, and R. S. Kroon, "Detecting potholes using simple image processing techniques and real-world footage," SATC, July 2015. Pretoria, South Africa.

[2] S. Nienaber, R. S. Kroon, and M. J. Booysen, "A comparison of low-cost monocular vision techniques for pothole distance estimation," in *IEEE CIVTS*, December 2015. Cape Town, South Africa.

[3] E. N. Ukhwah, E. M. Yuniarno, and Y. K. Suprapto, "Asphalt pavement pothole detection using deep learning method based on yolo neural network," in *2019 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pp. 35–40, IEEE, 2019.

[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.

[5] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, SenSys '08, (New York, NY, USA), p. 323–336, Association for Computing Machinery, 2008.

[6] G. D. D. Silva, R. S. Perera, C. Keppitiyagama, K. Zoysa, N. M. Laxaman, and K. Thilakarathna, "Automated pothole detection using wireless sensor motes," 2008.

[7] K. Bansal, K. Mittal, G. Ahuja, A. Singh, and S. S. Gill, "Deepbus: Machine learning based real time pothole detection system for smart transportation using iot," *Internet Technology Letters*, vol. 3, 03 2020.

[8] Aparna, Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, and A. Akula, "Convolutional neural networks based potholes detection using thermal imaging," *Journal of King Saud University - Computer and Information Sciences*, 2019.

[9] P. Ping, X. Yang, and Z. Gao, "A deep learning approach for street pothole detection," in *2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService)*, (Los Alamitos, CA, USA), pp. 198–204, IEEE Computer Society, aug 2020.

[10] J. Kim, J. Lee, K. Song, and Y.-S. Kim, "Vehicle model recognition using srgan for low-resolution vehicle images," in *Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition*, AIPR '19, (New York, NY, USA), p. 42–45, Association for Computing Machinery, 2019.

[11] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, pp. 740–755, Springer, 2014.

[17] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2472–2481, 2018.