BAE SYSTEMS

Deducing 15cm detail from 30cm satellite images with Deep Learning

By: Bingcai Zhang, Yen Luu, Kalyan Vaidyanathan, Fidel Paderes,
Reuben Settergren, and Kurt de Venecia

GXP®
Geospatial eXploitation Products™

Geospatial solutions
to ensure a safer world.

1. Executive summary

While the value of satellite imagery is highly dependent on image quality and resolution, these two components can now be enhanced through advanced Artificial Intelligence/Machine Learning (AI/ML) image processing techniques. Although algorithms used in image processing cannot create new information, we can use deep learning – a branch of AI – to create higher quality, higher resolution, Multispectral Images (MSI) from lower resolution satellite images, also known as super-resolution. A very successful and well-known example is the Zoom Video Communications, Inc. application, which uses deep learning to generate high quality and high resolution images without pushing a lot of pixels across the Web. Applying the same very deep Convolutional Neural Networks (CNN) in the geospatial domain (Kim et al., 2016; Shermeyer and Etten, 2019), we support manual and automatic imagery analysis workflows with superior-resolution satellite images.

Very deep CNN used in AI/ML workflows can learn the transformations between different zoom levels of image pyramids, also referred to as Resolution Sets (RSets). The CNN can learn the transformations from the 2:1 RSet at a Ground Sample Distance (GSD) of 60cm to the full resolution image at a GSD of 30cm by minimizing the differences between ground-truth full resolution and the derived 2x zoom. After training, the learned transformation is applied to the 1:1 full resolution image transforming the pixels to 2x resolution. The learned transformations, i.e., a CNN model, has intelligence built in and can infer higher resolution images.

Our case study indicates that the derivation of super-resolution images have the following advantages in imagery analysis workflows:

1. Significantly improved AI/ML object detection accuracy (positional accuracy, dimensional accuracy, orientation accuracy, precision, and recall).
2. Significantly improved manual feature extraction accuracy by allowing image analysts to place the extraction cursor at the precise feature edges and corners.
3. Significantly enhanced manual object/target search and identification workflow because of high quality and higher resolution images.
4. Enhanced image correlation for improvements in accuracy and quality in automatic terrain extraction for terrain analysis and navigation, and automatic tie point matching for triangulation and targeting.
5. Benefits to all imagery analysis workflows and derived products which use satellite images.

For our case study, we used a WorldView-4 image (courtesy of Maxar) of 30cm GSD panchromatic (PAN) and 120cm GSD MSI covering Tokyo. For automatic object detection with DeepObject™ (Zhang, 2016, 2017a, 2017b, 2020, 2021), the average relative positional accuracy (object center position) is about 15cm for super-resolution images vs. 30cm for panchromatic images. The precision has improved from less than 91.8% to 96.6% while the recall has also improved from 98.3% to 99.6%.

2. Deducing higher resolution pixels from lower resolution pixels

Resolving higher quality pixels from lower resolution pixels becomes feasible with deep learning. "Image processing cannot create new information" is a long-held truth in remote sensing and photogrammetry. Deep learning CNN can learn the transformations from lower resolution images to a higher resolution image for millions of different cases, and we can then use these learned transformations to deduce even higher resolution images.

The following explanation uses a simple example to illustrate how higher resolution pixels can be deduced from lower resolution pixels.

In **Figure 1**, there are 12 pixels of GSD 30cm. These 12 pixels represent a vertical step edge in the middle (pixel intensities stepping up from 54 to 80). Averaging 4 pixels into 1 pixel, we get 3 pixels of GSD 60cm (representing the 2:1 RSet). Given these three pixels of GSD 60cm, can we undo the information reduction and deduce the original 12 pixels of GSD 30cm? This is an ill-posed math problem and there is no single definitive solution. However, we use several iterations to demonstrate how to reliably interpret 12 pixels of GSD 30cm from 3 pixels of GSD 60cm. We use pixels with values 54 (p54) and 80 (p80) for this demonstration, which refer to the unknown values of the original 30cm pixels. The three pixels of GSD 60cm are denoted as p51, p67, and p83. The names '54' and '80' are just reminders of the ground-truth values we are trying to recover.

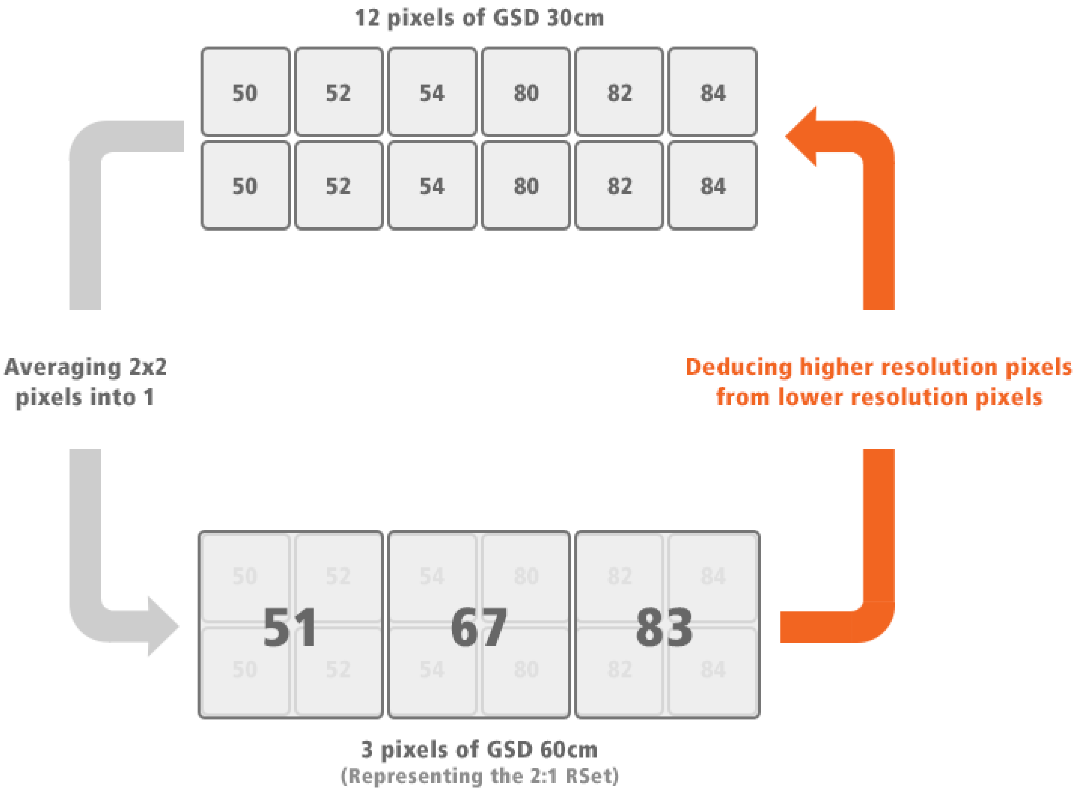


Figure 1*. From 30cm GSD image (above) to 60cm GSD image (below), we average 4 pixels into 1 pixel. In SOCET GXP®, we refer to the 30cm GSD as 1:1 RSet, and the 60cm GSD image as 2:1 RSet.*

In **Figure 2**, we split pixels of GSD 60cm into 4 pixels.

P54 is computed as:

p54 = w1 * p51 + w2 * p67 + w3 * p83, where w1, w2, and w3 are the weights for p51, p67, and p83 respectively. p80 = f1 * p51 + f2 * p67 + f3 * p83, where f1, f2, and f3 are the weights for p51, p67, and p83 respectively.

It should be noted that the weights for the same p51, p67, and p83 are different for p54 and p80. In the first iteration of CNN training, w1 and f1 are initialized to 0.0, w2 and f2 are initialized to 1.0, and w3 and f3 are initialized to 0.0. We get p54 = 67 = p80. The loss (the ground-truth value minus the inferred value) in both cases is +/-13.

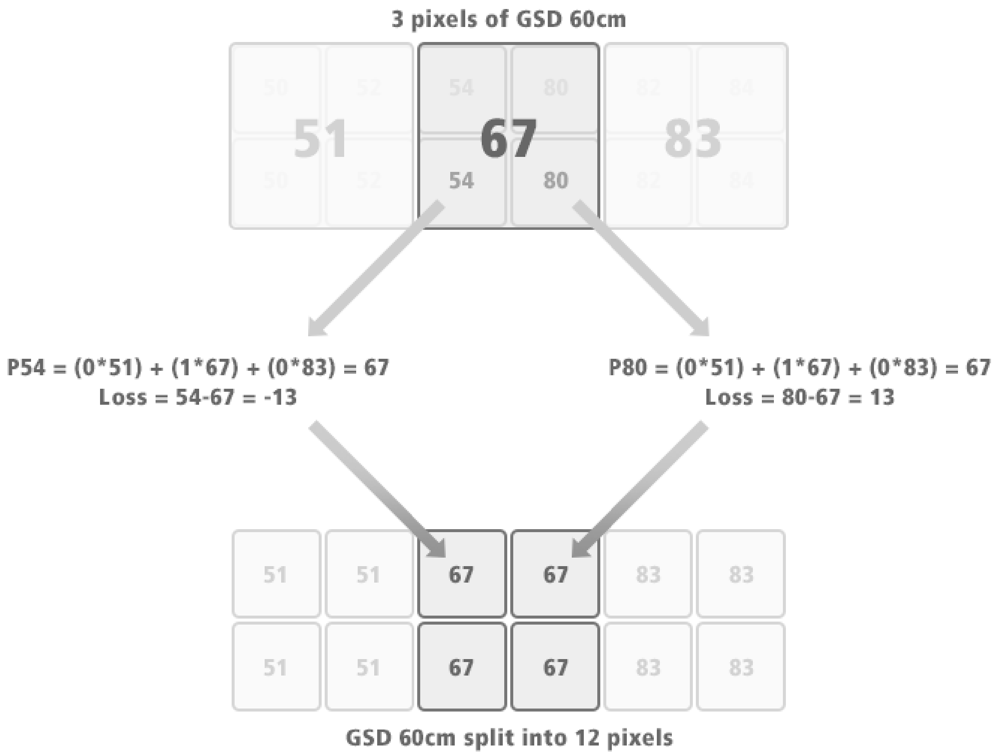


Figure 2*. Initially, we assign w1 and f1 to 0.0, w2 and f2 to 1.0, and w3 and f3 to 0.0. The deduced value p54 is 67 and p80 is 67. The losses are -13 and +13 respectively. In the next number of iterations, we adjust w1, f1, w2, f2, w3, and f3 such that the losses approach 0.0. The algorithms to adjust w1, f1, w2, f2, w3, and f3 based on losses are called gradient decent, which are the standard CNN algorithms.*

In **Figure 3**, we iteratively update w1, f1, w2, f2, w3, and f3 using gradient decent to minimise the losses. The losses may never converge to zero, which means we may never deduce a *perfect*, or 0-loss, higher resolution image, but CNN will enable us to achieve a *good*, or low-loss, recovery of the original image. The more training images we use, the closer we can get to perfectly recovering the original higher-resolution image.

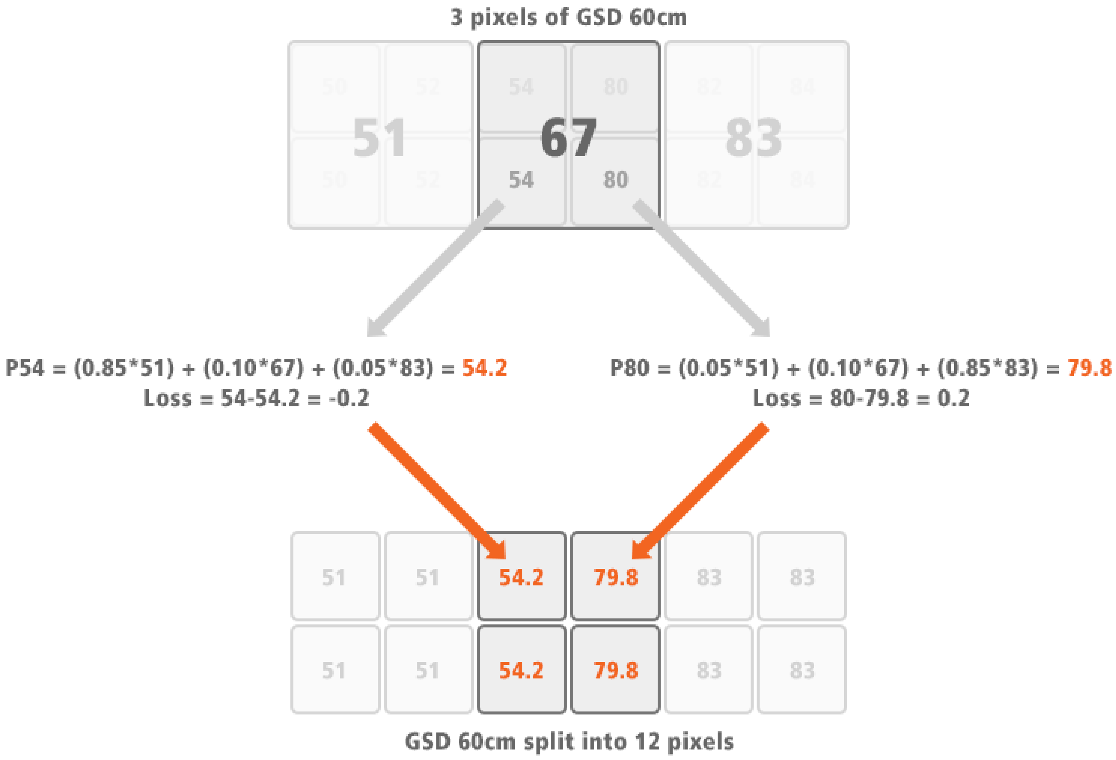


Figure 3. *Iteratively update weights to minimize losses. The losses may never converge to 0. To keep sharp vertical step edge, an experienced image analyst would split the edge pixel with the left half using the left pixel and the right half using the right pixel. In other words, image analysts split p67 by assigning p54 (left half) = 51 and p80 (right half) = 83. A CNN model trained by minimizing the loss function can learn the same higher-resolution behavior as a human image analyst.*

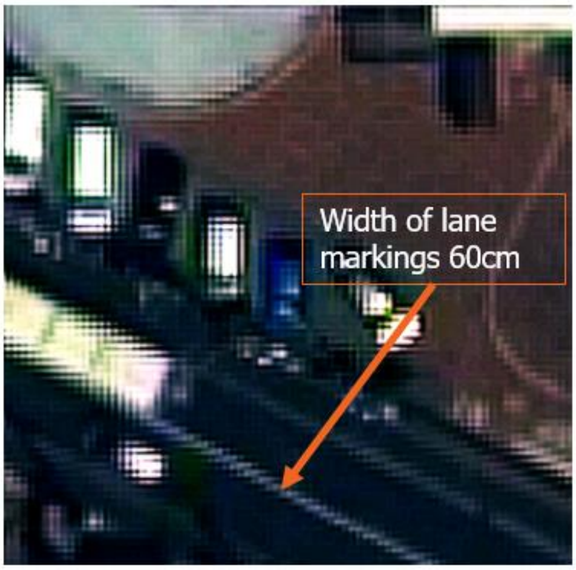
The CNN model learns the intelligence to preserve sharp vertical step image edges by minimizing the loss function. It's very generic and does not have the specific concept of "vertical step image edge". We can handcraft one algorithm for vertical step image edges, but we cannot handcraft one million algorithms to deal with one million other different cases. The CNN model learns all the transformations that occur between lower and higher resolution images, and then we can apply these learned transformations to our highest resolution imagery to deduce even higher resolution images (super-resolution). Although Figures 1, 2, and 3 are simple examples, they work in a similar fashion to the much more complex CNN model.

3. More accurate 3D measurements and more applications

Image analysts using imagery for site monitoring, or geospatial analysts doing foundation-mapping, benefit with more accurate 3-D measurements from super-resolution images. The super-resolution 15cm GSD MSI satellite imagery has more applications than the original 30cm GSD panchromatic images. In digital photogrammetry, accuracy and quality, also known as trusted GEOINT, are important factors when it comes to imagery exploitation and analysis. The resultant super-resolution imagery, which saves time while identifying objects in the area of interest, enhances the value of existing assets using the original-resolution satellite imagery. In **Figure 4**, the right image chip is from a super-resolution image of GSD 15cm. The left image chip is from the original resolution pan-sharpened image of 30cm GSD. In the right image, we can see circular edges (two rings) in the upper-left corner, but that detail is completely blurred in the left image. After trying several image enhancement algorithms on the left image, the lower image represents the best image enhancement that traditional algorithms can achieve. There is more information in the super-resolution image than the pan-sharpened, and enhanced lower image chips. How do we call this "more information"? If we call this new information, then, we are violating the long-held belief that "image processing cannot create new information". This 'new' information is not actually new, but rather **it is intelligence that resides in all of the training data. A CNN is able to learn this intelligence, and apply it to other data.**

Pan-sharpen image (30cm GSD)

Super-resolution image (15cm GSD)



Zoomed 60cm image (30 cm GSD)

Figure 4*. Super-resolution images can provide more accurate measurements and have a wider application. For instance, assume there was a traffic accident between a bus and a car. From the super-resolution image, we know that the bus was driving within its lane, while we cannot make the same determination from the pan-sharpened image, even after image enhancement. The width of the lane marking is measured at 28cm from super-resolution image vs. 60cm from pan-sharpened image. Image © 2021 Maxar Technologies.*

Super-resolution images make searching and recognizing man-made objects easier for image analysts. As shown in **Figure 5**, the traffic “STOP” sign painted on a road is much easier to recognize from a super-resolution image (15cm GSD, satellite MSI) than from a panchromatic image (30cm GSD). An image analyst may need human expertise to infer it must be a “STOP” sign from the panchromatic image, while it is easy to read “STOP” from the super-resolution image.

Figure 5*. Super-resolution images make searching and recognizing man-made objects easier. The right image is 15cm GSD, satellite MSI, deduced from the 30cm GSD pan-sharpen image on the left. The letters are much easier to recognize in the super-resolution image. CNN based super-resolution networks use AI to learn the transformations from lower resolution images to higher resolution images. Image © 2021 Maxar Technologies.*

4. Object detection accuracy comparison

Super-resolution images of 15cm GSD improves AI/ML object detection accuracy. We collected 1,755 positive training samples and 5,967 negative training samples from a WorldView-4 image covering Tokyo, Japan representing a total area over 201 square kilometers. We trained two DeepObject models: (1) using super-resolution 15cm GSD images; and, (2) using panchromatic 30cm GSD images. Our test site was an area where there were 526 densely parked cars (Figure 6). Because the cars were so densely packed together, AI/ML algorithms will generally have difficulty separating them, particularly when the gap between them might be one pixel or might not be seen at all due to the obliquity in the image collection causing the ground between the cars to be occluded. As a result, when using the original panchromatic images with DeepObject, the precision of the results was 91.8%. When running the same test case using the super-resolution images (15cm GSD) instead of the original panchromatic images (30cm GSD), DeepObject achieved a precision of 96.6%.

	True positive	False positive	Recall	Precision	Average positional errors
Panchromatic (30cm GSD)	517	46	98.3%	91.8%	30cm
Super-resolution (15cm GSD)	524	18	99.6%	96.6%	15cm

***Table 1**. **Object detection accuracy comparison**. With the same training samples of 526 small vehicles, we trained two models: (1) with super-resolution images of GSD 15 cm; and, (2) with panchromatic images of GSD 30cm. We discovered the recall results, which are defined as true_positive/total_objects, and the precision results, which are defined as true_positive/(true_positive + false_positive)" have both significantly improved.*

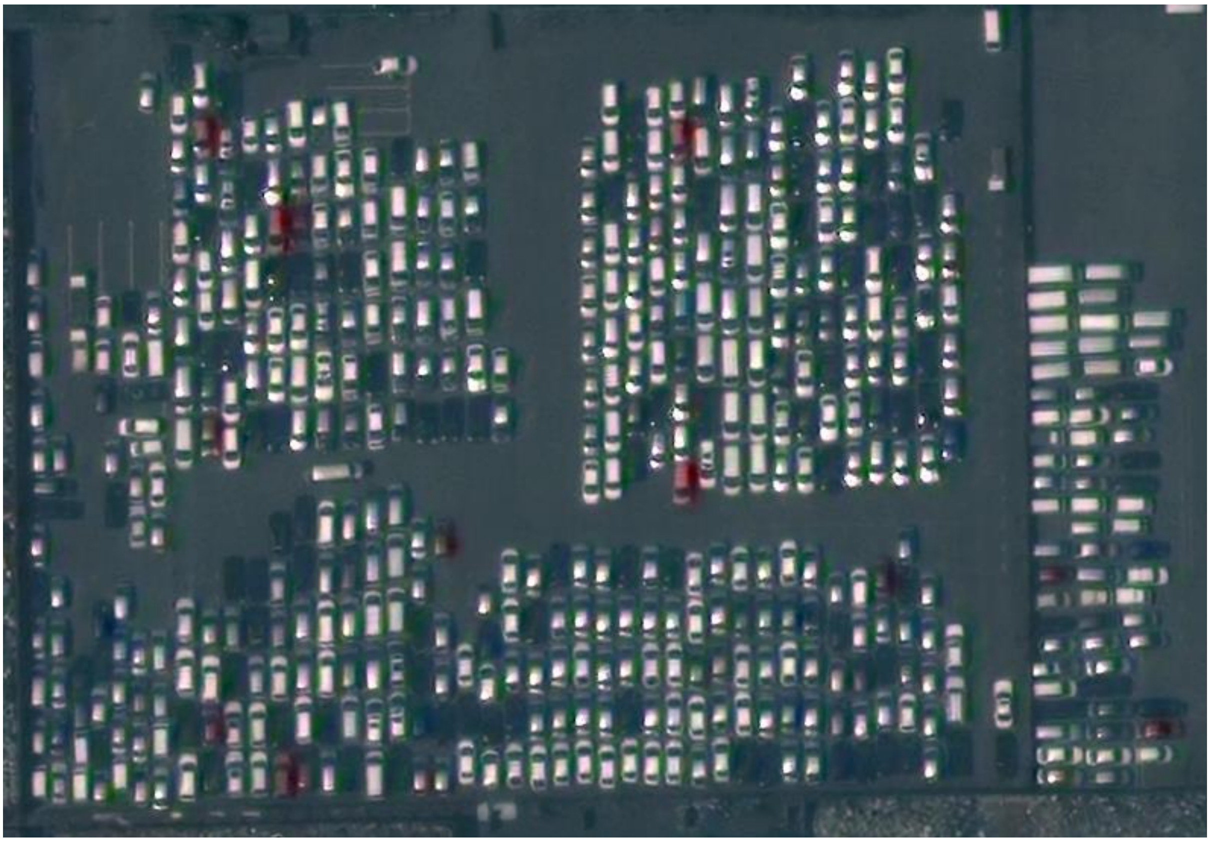


***Figure 6**. **Object detection accuracy evaluation site**. Because the 526 small vehicles in this area are densely parked, it is challenging for automatic object detection to individually identify these vehicles. Therefore, DeepObject was used for the accuracy evaluation. Image © 2021 Maxar Technologies.*

***Figure 7**. Object detection accuracy in an original panchromatic image of 30cm GSD. There are nine false negatives (or missing detections, yellow boxes) and **more than 46 false positives** (red boxes). The recall is 98.3% and the precision is less than 91.8%. Many bounding boxes are not precisely covering vehicles because the average positional accuracy is **30cm**. For example, when two bounding boxes cover the same vehicle, we do not count one of them as false positive. As a result, the precision is actually lower than 91.8%. Image © 2021 Maxar Technologies.*

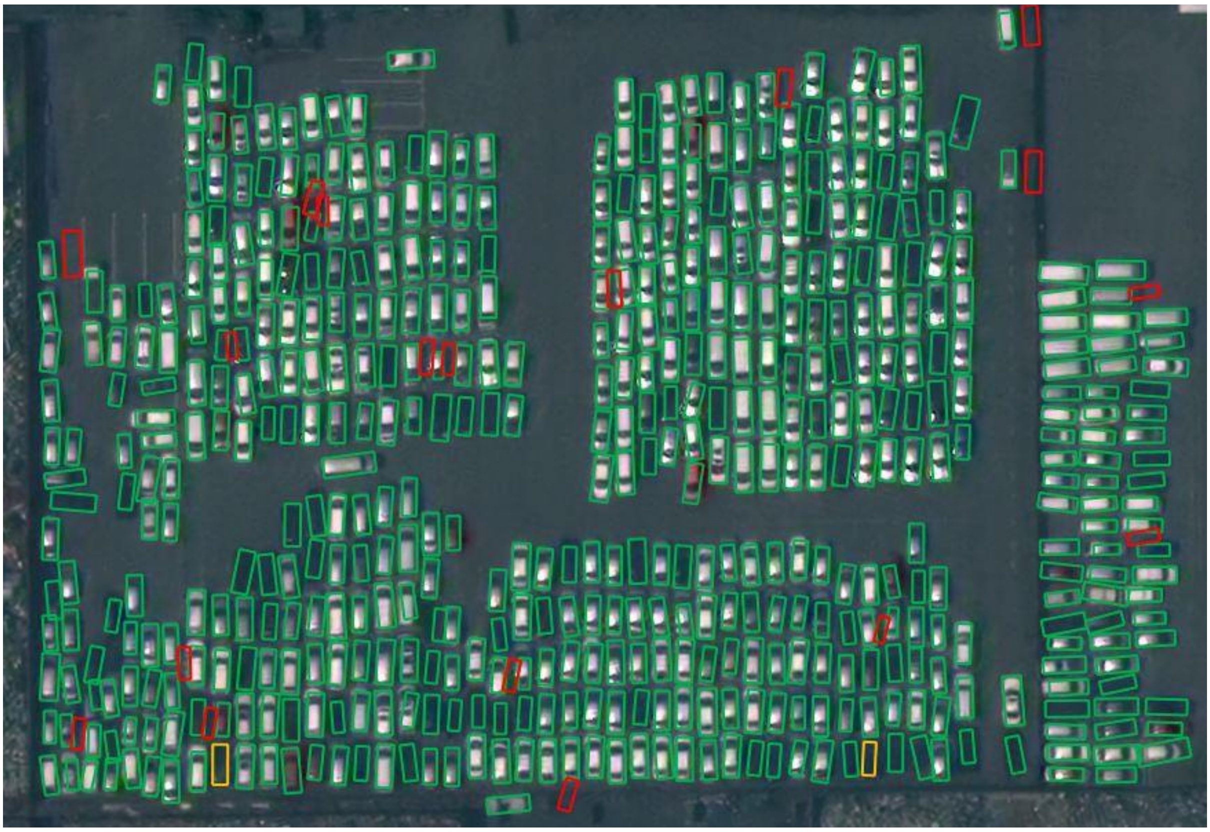


Figure 8. *Object detection accuracy in a super-resolution image of 15cm GSD. There are two false negatives and 18 false positives. The two missing detections are black vehicles. The recall is 99.6% and the precision is 96.6%. Bounding boxes cover vehicles with much higher positional accuracy vs. bounding boxes in Figure 7, because the average positional accuracy is 15cm. Image © 2021 Maxar Technologies.*

Object detection with super-resolution images have significantly higher precision, recall, as well as higher positional accuracy. DeepObject has a specific model to estimate object centers. With a super-resolution image, the average center estimation error is about 15cm vs. 30cm for panchromatic images. The bounding boxes in **Figure 8** vs. in **Figure 7** show that:

1. Super-resolution images can achieve higher positional accuracy for automatic object detection.
2. Super-resolution images can separate objects, which are very close to each other, for automatic object detection.
3. Super-resolution images can achieve significantly higher object detection precision than panchromatic images.
4. Super-resolution images can achieve higher accuracy of estimating object lengths and object orientation angles. Bounding boxes in **Figure 8** fit more precisely to vehicles than those in **Figure 7**.

5. Conclusion

Super-resolution based on very deep CNNs learns the transformations from lower resolution images to higher resolution images. These learned transformations, also known as AI, can be used to create higher resolution images from original imagery. Super-resolution 15cm GSD, satellite MSI enhance the value of original 30cm GSD satellite images. With higher resolution, the images become relevant for more applications allowing image analysts to search and find man-made objects more easily.

Super-resolution 15cm GSD satellite images can achieve higher measurement accuracy for manual feature extraction used in foundation mapping and 3-D modelling applications by resolving and using image edges and corners of man-made features. Since super-resolution images preserve image edges accurately, the manually extracted features have accordingly higher accuracy.

Super-resolution satellite images can achieve much higher positional accuracy (object center positions and object orientation angles), recall, and precision than panchromatic satellite images for automatic object detection. Our case study found that the average relative positional accuracy (center position) is about 15cm for super-resolution images vs. 30cm for panchromatic images. Super-resolution images can achieve higher accuracy of object dimensions (object length) than panchromatic images when using DeepObject.

6. Acknowledgments

We want to thank our co-workers: Radoslav Gaidadjiev, Shannon McDonald, Brian Lei, Mikaela Peck, and Gianni LaChica for their supports.

7. Bibliography

Kim, J., J. K. Lee and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646-1654, doi: 10.1109/CVPR.2016.182.

Shermeyer, J., Etten, A. V., CosmiQ Works, In-Q-Tel "The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery", 2019, Pages: 1432-1441, DOI: 10.1109/CVPRW.2019.00184

Zhang, B., 2016. Big geospatial data + deep learning + high performance computing = geospatial intelligence, presentation S6260, 2016 GPU Technology Conference, San Jose, CA, 4-7 April. [http://on-demand.gputechconf.com/gtc/2016/presentation/s6260-bingcai-zhang-big-geospatial-data-deep-learning-high-performance-computing.pdf]

Zhang, B., 2017a. 3D DeepObject for precision 3D mapping, featured presentation S7149, 2017 GPU Technology Conference, San Jose, CA, 8-11 May. [https://on-demand.gputechconf.com/gtc/2017/presentation/s7149-bingcai_zhang-deepobject-for-precision-printing.pdf]

Zhang, B. 2017b. Rotation variant object detection in deep learning, Nov. 8. US Patent #10347720 [https://patents.google.com/patent/US10346720B2/en]

Zhang, B. and R. Hammoud, 2019. Fast and accurate target detection in overhead imagery using double convolution neural networks, invited presentation, SPIE Defence + Commercial Sensing, Baltimore Convention Center, Baltimore, Maryland, 14-18 April. [https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10988/109880O/Fast-and-accurate-target-detection-in-overhead-imagery-using-double/10.1117/12.2519795.full?SSO=1]

Zhang, B. 2021. Intelligent photogrammetry for digital elevation model production. SPIE Defence + Commercial Sensing, 12-16 April. [https://www.spie.org/Publications/Proceedings/Paper/10.1117/12.2595802?SSO=1]

BAE Systems, Inc. is the U.S. subsidiary of BAE Systems plc, an international defense, security and aerospace company which delivers a full range of products and services for air, land, and naval forces, as well as advanced electronics, security, information technology solutions, and customer support services.

Enabling development of the most advanced geospatial intelligence, BAE Systems GXP™ software enables rapid discovery, exploitation, and dissemination of mission-critical geospatial data. From key military, security, and incident response operations, to a variety of commercial development and research initiatives, GXP provides a comprehensive suite of solutions to inform effective decision-making and ensure a safer world.