Unsupervised Deep Learning Parameter Estimation for High Fidelity Synthetic Aperture Radar Super Resolution

Matthew Tay
Imaging Radar Laboratory (Sensors Division)
DSO National Laboratories
Singapore
thanyang@dso.org.sg

***Abstract*—Super-Resolution for Synthetic Aperture Radar (SAR) has been of high interest for its applications in reducing system Size Weight and Power (SWAP) and potential to improve SAR interpretability. Yet to enable super-resolution techniques like Super Spatially Variant Apodization (SSVA), precise knowledge of the Synthetic Aperture Radar (SAR) imaging parameters is required. Such parameters include the pixel, ground range and azimuth resolution which may be missing or tedious to parse. Moreover, these parameters are dependent on terrain relief and the effective target beam-width which may be absent in practical scenarios involving real targets. The inability to calculate sampling parameters would degrade side-lobe levels and image quality in super-resolution processing. To tackle this, we design a novel deep learning network that leverages a bi-linear sampling layer and a total variation loss that is able to directly estimate the required sampling factor without supervision. Our network is able to learn and carry out optimal side-lobe cancellation with no prior knowledge of the system or target imaging parameters. In doing so, we make it possible to carry out super-resolution on SAR images with no a-prior expert knowledge and outperform traditional algorithms in presence of imperfect expert knowledge. We validate our approach on both simulated data and a dataset (GOTCHA) collected by Airforce Research Laboratory (AFRL).**

***Index Terms*—Unsupervised Deep Learning, Super-resolution, Signal Processing, Sidelobe cancellation**

I. Introduction

Traditionally, machine learning for Synthetic Aperture Radar (SAR) have explored mostly applications in target classification [1] [2]. Alongside rapid developments in the field, researchers have explored utilizing deep learning for more applications traditionally handled by signal processing, including image formation, de-noising and super-resolution [3] [4] [5]. Most of these applications have one of a few characteristics: 1) Training is done via availability of ground truth training data, 2)is difficult to generalize to new data, 3)has difficulty with interpretation beyond a 'black box' or 4)requires expert knowledge in SAR.

To address the above, we align components of our neural network operations with data processing flows known to produce theoretically optimal results in signal processing. In addition, outputs from our network produce physically interpretable quantities. We are able to do this by taking advantage of the simplicity of mapping raised cosine tapers in the frequency domain to 3 point convolutions in the image domain [6]. With a suitable cost function measuring image focus, the resulting network can compute physically interpretable optimal parameters for side-lobe cancellation. Such cancellation is central for super-resolution SAR. Thus the resulting network termed Deep Learning Super Resolution (DLSR) enables super-resolution with no aprior expert knowledge on the image formation parameters and no annotated ground truth.

II. Prior Work

A. Signal processing

SAR Super resolution has been first proposed by Stankwitz [6] in 1996 and termed Super Spatially Variant Apodization (SSVA) as an extension to the optimal sidelobe cancellation algorithm of spatially variant apodization [7]. In comparison to Minimum Variance [18] Beamforming methods, SSVA is computationally tractable by avoiding covariance matrix inversions. However, the downside of SSVA is that it requires data sampled at an integer multiple of the nyquist rate. Later, authors like Smith [8] and later in 2015 Xiong et al [9] extended the method to non-integer nyquist sampled data. We further extend these works, showing that critical parameters like the sampling rate can be directly estimated from the data.

B. Computer vision and machine learning

Super resolution for electro-optical images has been extensively studied by authors like Dong et al [10]. Recently, work by Li et al [11] have adopted similar networks for SAR images. These networks adopt an underlying training strategy of taking a high resolution image x, degrading it by some noise y, and training the neural network to recover x or y from training sample pairs of (x, x+y). In our work, we prefer to train in an unsupervised manner because 1)there is a question of why the network is needed when high resolution data is available and 2)networks trained via supervised learning are known to have difficulty generalizing to new data-sets with

different pixel distributions. Recently, work by Wei et al [12] adopt an approach closest to our work. However, their work is based on a compressive sensing framework to SAR imaging, instead of optimal apodization theory here. In addition, their work is based on imaging and requires access to the SAR system parameters like transmit frequency, while our work leverages the simplicity of apodization theory does not require such parameters.

III. Method

A. Optimal Apodization

We start with a recap of the ideal SAR impulse response, which can be described as a sinc function. In one dimension, the impulse response x for pixel index m of total number of pixels M spanning a spatial extent of M_{extent} in meters may be described as a function of its oversampling parameter $1/W_s$ by:

$$x[m] = sinc(W_s m(\frac{M_{extent}}{M})) \tag{1}$$

Where $sinc(x) = \frac{sin(\pi x)}{\pi x}$. The sampling factor W_s may be calculated by the ratio of the sampled resolution ρ_{pixel} to the actual resolution ρ_{actual}:

$$W_s = \frac{\rho_{pixel}}{\rho_{actual}} \tag{2}$$

In addition, in azimuth prior to azimuth compression the sampling factor has the added interpretation of being the ratio of the length of non-zero transmit pulses K to the total length of zero padded data N:

$$W_s = \frac{K}{N} \tag{3}$$

In two dimensions x and y, the two dimensional impulse response s indexed by pixel indices m and n may be similarly described as the product of 2 one dimensional impulse responses $x[m], y[n]$:

$$s[m,n] = x[m]y[n] \tag{4}$$

With no loss of generality, we continue the discussion in one dimension for notation clarity. It is known that due to finite aperture or transmit bandwidth, a point target post SAR compression creates a ringing artifact known as side-lobes. Side lobes are an unwanted addition to SAR images, because they create artifacts and can be misinterpreted as or mask real targets. To address this, it is common to apply a frequency domain weighting like the hamming or hanning window to the data before compression. In particular, a large class of windows like the hamming window in the frequency domain are described by a raised cosine taper. For a filter frequency response indexed by k in the frequency domain of total number of samples N, the tapers are described as a function of a weight w:

$$filter(k) = 1 + 2wcos(\frac{2\pi k}{N}) \tag{5}$$

At varying levels of w, the sidelobe suppression properties of the taper window changes, and $w = 0$ corresponds to no taper while $w = 0.5$ corresponds to the hamming window. Figure 1 shows an illustration of these tapers, However, the downside to these windows is that there is a resolution degradation. For example, the hamming window has a resolution broadening factor of 1.3.

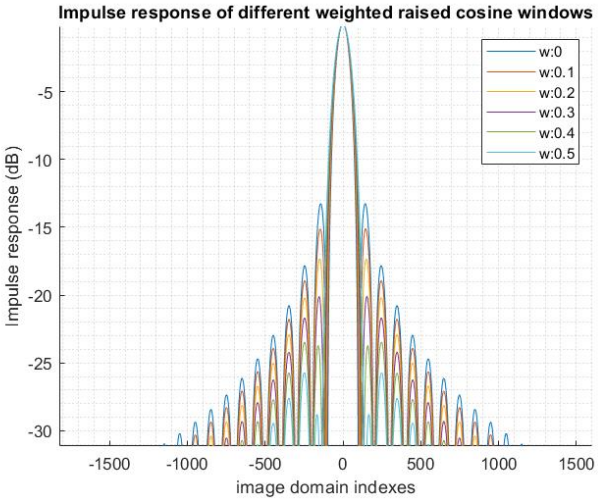

Fig. 1. Impulse response of different raised cosine tapers parameterized by weighting factor w.

On the other hand, as Stankwitz [6], Smith [8] and Xiong [9] has shown, the problem of minimizing the sidelobes may be reframed as minimizing total signal intensity while preserving the gain at mainlobe and monotonicity constraints. Figure 2 show such an operation.

$$\begin{aligned} \min_{w} \quad & |x[m] * IDFT(filter(w))[m]| \\ \text{st} \quad & filter[0] == 1 \\ & filter[0] > filter[K/2-1] \\ & filter[K/2] > 0 \end{aligned} \tag{6}$$

While seemingly complicated, the IDFT of the cosine function is simply the dirac delta function. Thus the optimal kernel is:

$$w[m] = a\delta[m] + w(\delta[m - \frac{1}{W_s}] + \delta[m + \frac{1}{W_s}]) \tag{7}$$

After convolution with the signal letting the derivative w.r.t the weight w be 0, the optimal weight is:

$$\begin{aligned} w[m] = & -\frac{x[m]}{x[m-\frac{1}{W_s}] + x[m+\frac{1}{W_s}]} \\ \text{st} \quad & \frac{1}{2} >= w >= 0 \end{aligned} \tag{8}$$

It can be noted that the apodization is carrying out a weighting of the original signal with a linear addition of the sum of its neighboring nyquist sampled neighbors. Since the optimal weights can only vary between 0 and $\frac{1}{2}$, we can simply compute the weighted pixel value $x'[m]$ optimal weight at $w = \frac{1}{2}$:

$$x'[m] = x[m] + w(x[m - \frac{1}{W_s}] + x[m + \frac{1}{W_s}]) \tag{9}$$

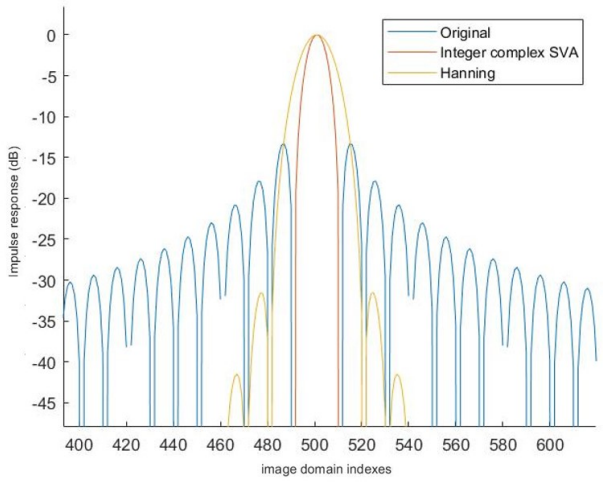

Fig. 2. Illustration of optimal sidelobe suppression while preserving resolution via Spatially Variant Appodization SVA for complex valued SAR data.

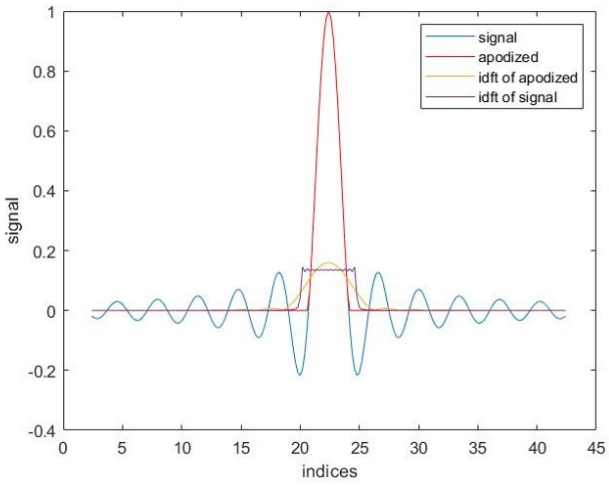

Fig. 3. Illustration of increased bandwidth after apodization

If there is a sign change from the original signal, we can infer that an optimal apodization is possible, and we can set the output image pixel to 0. If no sign change occurs, we can set the image pixel to minimum of the original pixel and the weighted pixel. The above is an efficient implementation of the following calculation.

$$x'[m] = \begin{cases} x[m] & w <= 0 \\ 0 & 0 < w <= \frac{1}{2} \\ x[m] + \frac{1}{2}(x[m - \frac{1}{W_s}] + x[m + \frac{1}{W_s}]) & w > \frac{1}{2} \end{cases} \tag{10}$$

With an apodized signal, an interesting change occurs in the frequency domain. As seen in Figure 3, the bandwidth in the frequency domain has increased. Thus one can reweight the spectrum and obtain increased resolution.

B. Machine Learning

However, often the sampling information $\frac{1}{W_s}$ may not be well known to the users of the SAR image before hand. This is the scenario for example, when using images without access to terrain elevation. Alternatively, expert information about the azimuth resolution may be degraded due to the target beam-width limiting the actual effective integrated beam-width. With imperfect or missing information, the apodization procedure will be degraded, leading to suboptimal super-resolution computation.

Thus we design a self-supervised network to address these concerns and estimate underlying sampling rate. First, two subnetworks that are each made of:

- Convolutional Kernel with 8x8 filter size, 5 filters
- Rectified Linear Unit
- Maximum Pooling with 2x2 pooling size
- Convolutional Kernel with 8x8 filter size, 5 filters
- Rectified Linear Unit
- Fully Connected layer with 32 output size
- Fully Connected layer with a single output

are defined, produces a single subpixel estimate of the shift $\frac{1}{W_s}$ in each of the 2 image axes.

Next, the shifts are input to a bilinear sampling layer [17] that samples the corresponding image samples $x[m + \frac{1}{W_s}]$ and $x[m - \frac{1}{W_s}]$ for each corresponding input sample $x[m]$. Given the values of the input to sample from (x_i^s, y_i^s) that are in general floating points, the sampling kernel is defined as follows to map the output V_i from the input U. The image is indexed by the (n, m) and the channel of interest c.

$$V_i^c = \sum_n \sum_m U_{nm}^c max(0, 1 - |x_i^s - m|) max(0, 1 - |y_i^s - n|) \tag{11}$$

The $1 - |x_i^s - m|$ weights pixels inversely proportional to their distance from the sampling point. The layer is differentiable with derivatives defined [17]:

$$\frac{\partial V_i^c}{\partial x_i^s} = \sum_n \sum_m U_{nm}^c max(0, 1 - |y_i^s - n|) \begin{cases} 0, & |m - x_i| > 1 \\ 1, & m > x_i^s \\ -1, & m < x_i^s \end{cases} \tag{12}$$

$$\frac{\partial V_i^c}{\partial U_{nm}^c} = max(0, 1 - |x_i^s - m|) max(0, 1 - |y_i^s - n|) \tag{13}$$

Finally, the sampled layers may be passed to an apodization layer implementing (10).

The network is defined by a cost function penalizing the total variation loss of the image, which can be defined [13] as:

$$V_{loss} = \sum_n \sum_m |V[n+1, m] - V[n, m]| + |V[n, m+1] - V[n, n]| \tag{14}$$

IV. Experimental Data

A. Simulated Data

We simulate an impulse response created by 4 closely spaced scatterers, as seen in Figure 4. The scatterers are

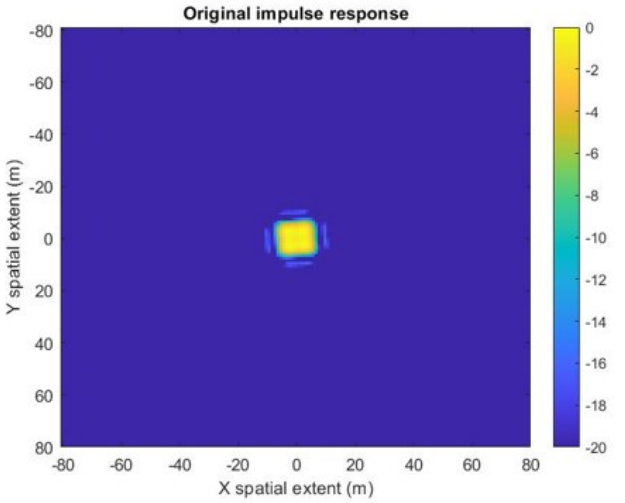

Fig. 4. Simulated SAR image of closely spaced scatterers. Scatterers not distinguishable clearly.

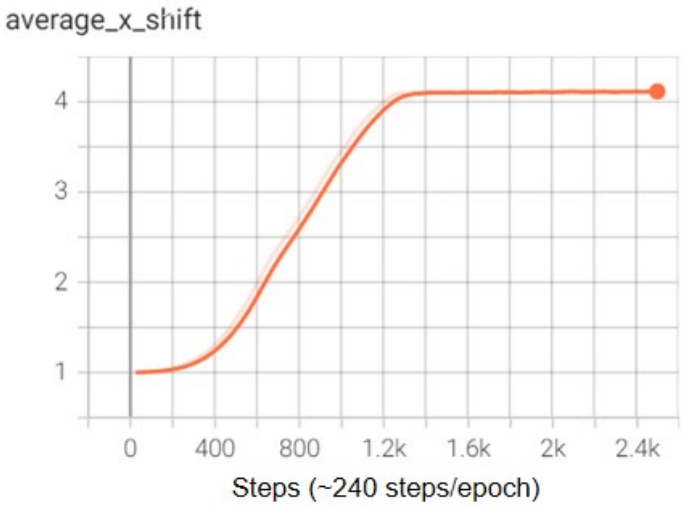

Fig. 5. Neural network estimating sampling rate shift without supervision. Ground truth is 4 pixels.

not easily distinguishable from each other. Use an ADAM optimizer with a learning rate of 5e-4 and running our network for 10 epochs, we quickly converge to output sample rate of $\approx 4.1 - 4.2$ pixels. This may be observed in Figure 5 and a similar plot is obtained for the y-axis.The ground truth is 4 pixels. The sub-pixel accuracy of the network estimation might be improved with further tuning of the learning rates and may be subject to future study.

In Figure 6 we see that with just a blind estimate of 1 for the sampling rate, a super-resolution algorithm fails to clearly focus the individual scatterers. On the other hand, when we add the network estimated sampling rate with no supervision, the scatterers can be resolved as seen in Figure 7.

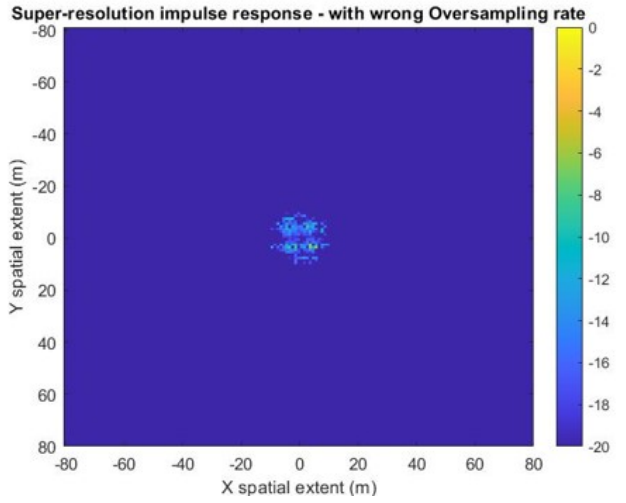

Fig. 6. With a blind estimate of 1 pixel sample spacing, super-resolution algorithm on its own is unable to super-resolve scatterers.

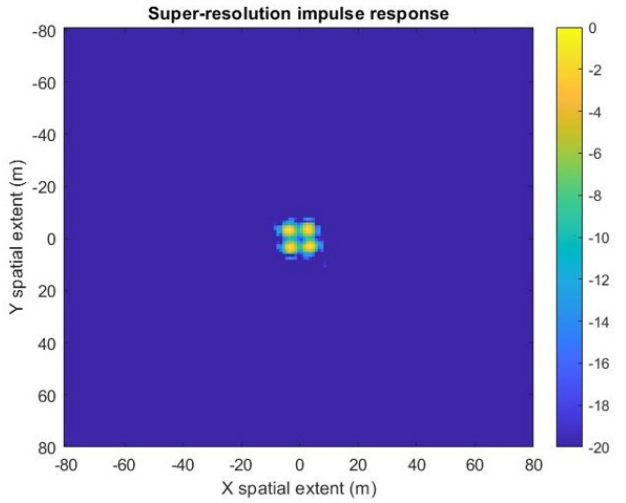

Fig. 7. 4 scatterers clearly resolved after an unsupervised neural network estimate of sampling rate is passed to super-resolution algorithm.

B. Real Data

The GOTCHA volumetric dataset [14] is an X-band collection of phase history of a parking lot. In the scene, there are a few calibration targets, as well as cars parked in the parking lot. The transmit bandwidth is approximately 640 MHz, and the flight is travelling in a circular manner. 3 degrees of azimuth beam width is integrated for the displayed image in Figure 7. As noted in signal processing literature [16] the azimuth resolution is defined based on the broadening factor K_a, the carrier frequency λ_c and the integration angle $\Delta\theta$:

$$\rho_{az} = \frac{\lambda_c K_a}{2\Delta\theta} \tag{15}$$

In particular, the broadening factor K_a is introduced by processor aperture weighting for side-lobe control, and by residual amplitude and phase errors. It was not clearly mentioned

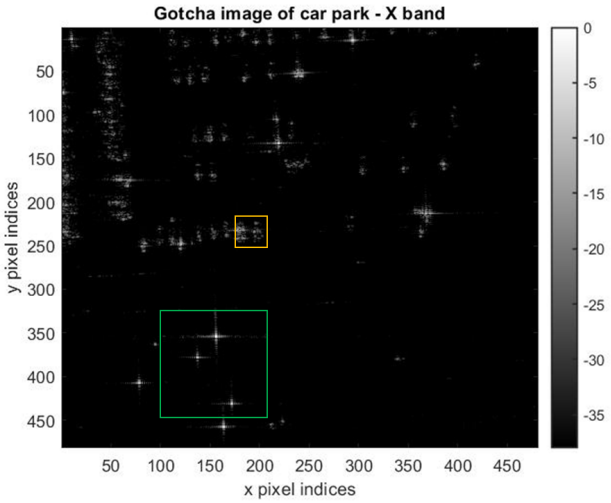

Fig. 8. Gotcha X-band image of car-park lot. 2 vehicles parked side by side (Orange square) and calibration targets (Green square) are cropped for further processing.

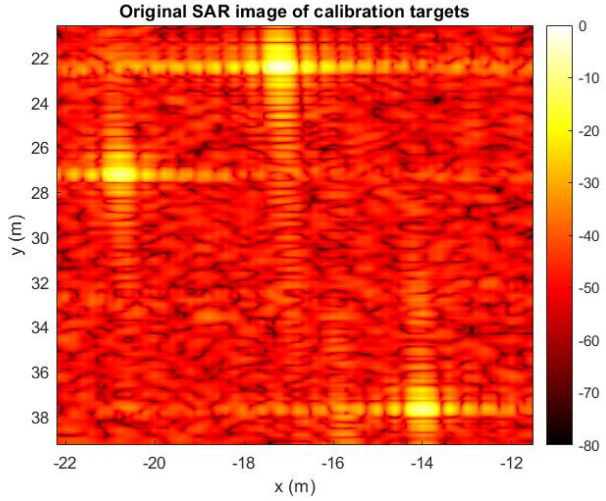

Fig. 9. Calibration targets prior to super-resolution.

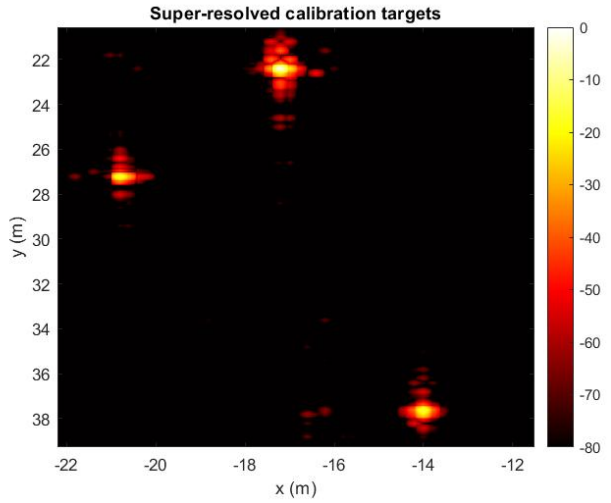

Fig. 10. Calibration targets after to super-resolution with some assistance on parameter estimation from unsupervised deep neural network.

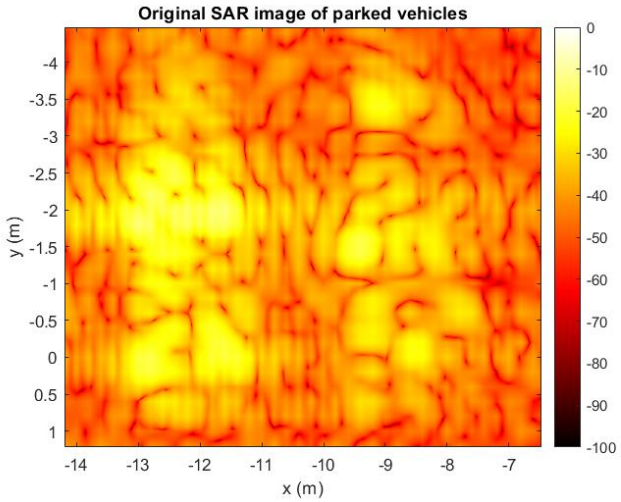

Fig. 11. 2 parked vehicles prior to super-resolution.

whether the processor had applied weighting, and as such there is an ambiguity for the optimal parameter to use.

We pass the cropped image to our network, which learns an optimal shift distance $\frac{1}{W_s}$ of approximately 2 in both azimuth and range directions. The resulting estimates may be passed to the super resolution processing algorithm and the results are seen in Figure 9 and Figure 10 respectively for corner reflector targets and in Figure 11 and Figure 12 for the parked vehicles.

We next test the scenario where a person blindly uses a rectangular weighting of $K_a = 0.89$ in azimuth. This might be based on the wrong assumption that no amplitude and phase errors in azimuth occurred, and that a rectangular aperture broadening for a real aperture is used. With 3 degrees of integration, this yields a pixel shift of 1.4 instead of 2 estimated by the network. We plot the azimuth cuts of the calibration target in Figure 13 showing the superior apodization performance (peak amplitude close to original, and no sidelobe degradation) when the neural network estimated shift is used.

Finally, we plot the resolution enhancement in range and azimuth of super-sva after assistance from DLSR neural network estimates, further validating the processing pipeline. The range and azimuth cuts of the calibration target is shown in Figure 14 and summarized in Table I. Typical resolution enhancement of 40-60% is reported in literature. It should be noted that the improvement in resolution is dependent on the shape of the apodization and is not within current scope of the paper.

V. Conclusion and Discussion

In this work, we show that it is possible to estimate parameters from a SAR image without any annotated ground truth, and in doing so enable super-resolution with no apriori knowledge of the scene, nor of the system. We base our work on the observation that optimal apodization can be

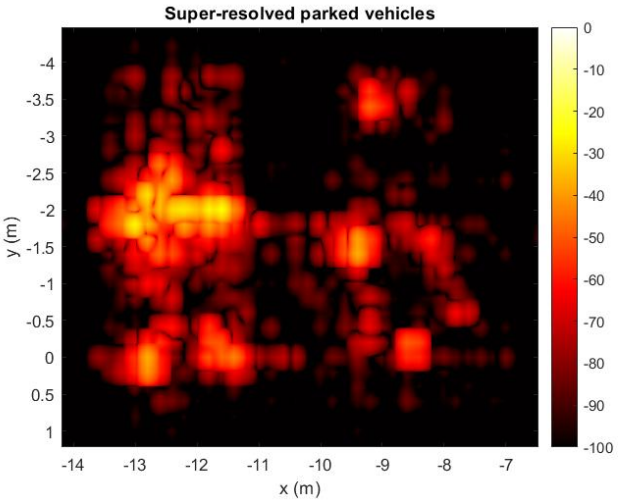

Fig. 12. 2 parked vehicles after super-resolution with some assistance on parameter estimation from unsupervised deep neural network.

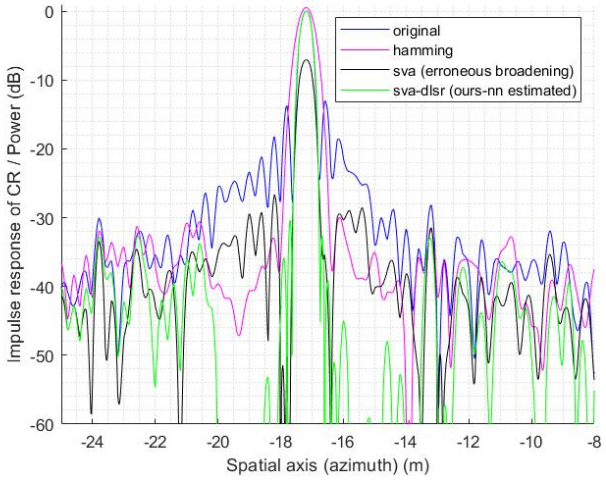

Fig. 13. Azimuth cut of calibration target of: Original image (blue), Hamming weighted (pink), Spatially Variant Apodization with wrong pixel shift parameter (black) and Spatially Variant Apodization with neural network assisted parameter estimation (green).

represented with well defined derivatives via bilinear sampling in a neural network, and that the total variation loss can capture ringing artifacts induced by presence of side-lobes. The current approach has characterized the performance for point scatterers. Future endeavours can explore enhancing and characterizing imaging performance for distributed targets.

TABLE I
RESOLUTION ENHANCEMENT PERFORMANCE

Metric	**Configuration**		
	Original	***DLSR+SVA***	***Improvement %***
Range Resolution	0.31m	0.18m	42
Azimuth Resolution	0.38m	0.18m	53

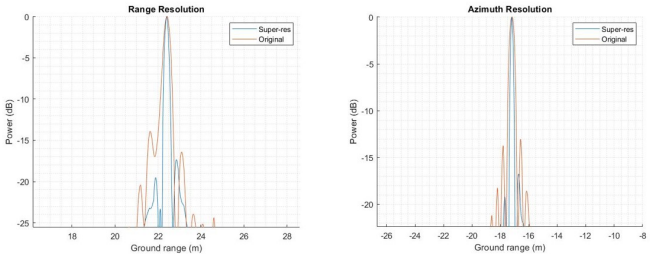

Fig. 14. Range (Left) and Azimuth (Right) cuts of calibration target after unsupervised neural network assisted super-resolution (Blue). Original impulse response of calibration target is plotted in (Red).

ACKNOWLEDGMENT

The author would like to thank DSO National Laboratories for supporting this work.

REFERENCES

[1] S. Chen and H. Wang, "SAR target recognition based on deep learning," 2014 International Conference on Data Science and Advanced Analytics (DSAA), 2014, pp. 541-547, doi: 10.1109/DSAA.2014.7058124.

[2] C. Coman and R. Thaens, "A Deep Learning SAR Target Classification Experiment on MSTAR Dataset," 2018 19th International Radar Symposium (IRS), 2018, pp. 1-6, doi: 10.23919/IRS.2018.8448048.

[3] J. Gao, B. Deng, Y. Qin, H. Wang and X. Li, "Enhanced Radar Imaging Using a Complex-Valued Convolutional Neural Network," in IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 1, pp. 35-39, Jan. 2019, doi: 10.1109/LGRS.2018.2866567.

[4] P. Wang, H. Zhang and V. M. Patel, "SAR Image Despeckling Using a Convolutional Neural Network," in IEEE Signal Processing Letters, vol. 24, no. 12, pp. 1763-1767, Dec. 2017, doi: 10.1109/LSP.2017.2758203.

[5] L. Lin, J. Li, Q. Yuan and H. Shen, "Polarimetric SAR Image Super-Resolution VIA Deep Convolutional Neural Network," IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 3205-3208, doi: 10.1109/IGARSS.2019.8898160.

[6] H. C. Stankwitz and M. R. Kosek, "Sparse aperture fill for SAR using super-SVA," Proceedings of the 1996 IEEE National Radar Conference, 1996, pp. 70-75, doi: 10.1109/NRC.1996.510659.

[7] H. C. Stankwitz, R. J. Dallaire and J. R. Fienup, "Spatially variant apodization for sidelobe control in SAR imagery," Proceedings of 1994 IEEE National Radar Conference, 1994, pp. 132-137, doi: 10.1109/NRC.1994.328113.

[8] Smith BH. Generalization of spatially variant apodization to noninteger Nyquist sampling rates. IEEE Trans Image Process. 2000;9(6):1088-93. doi: 10.1109/83.846250. PMID: 18255478.

[9] T. Xiong, S. Wang, B. Hou, Y. Wang and H. Liu, "A Resample-Based SVA Algorithm for Sidelobe Reduction of SAR/ISAR Imagery With Noninteger Nyquist Sampling Rate," in IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 2, pp. 1016-1028, Feb. 2015, doi: 10.1109/TGRS.2014.2332252.

[10] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Image Super-Resolution Using Deep Convolutional Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Preprint, 2015

[11] Yanshan LI, Li ZHOU, Fan XU, Shifu CHEN, OGSRN: Optical-guided super-resolution network for SAR image, Chinese Journal of Aeronautics, 2021

[12] Y. Wei, Y. Li, Z. Ding, Y. Wang, T. Zeng and T. Long, "SAR Parametric Super-Resolution Image Reconstruction Methods Based on ADMM and Deep Neural Network," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 12, pp.10197-10212, Dec. 2021, doi: 10.1109/TGRS.2021.3052793.

[13] "Total variation denoising," Wikipedia, 22-Feb-2022. [Online]. Available:https://en.wikipedia.org/wiki/Total_variation_denoising. [Accessed: 01-Mar-2022].

[14] "Gotcha Volumetric SAR Dataset," Gotcha volumetric SAR data set overview. [Online]. Available:https://www.sdms.afrl.af.mil/index.php?collection=gotcha.[Accessed: 01-Dec-2021].

[15] A. W. Doerry, “Catalog of window taper functions for sidelobe control,” Catalog of Window Taper Functions for Sidelobe Control (Technical Report) — OSTI.GOV, 01-Apr-2017. [Online]. Available: https://www.osti.gov/biblio/1365510. [Accessed: 07-Apr-2022].

[16] Walter G. Carrara, Ron S. Goodman and Ronald M. Majewski, ”Spotlight Synthetic Aperture Radar Signal Processing Algorithms,” Artech House, 1995.

[17] Jaderberg, Max Simonyan, Karen Zisserman, Andrew Kavukcuoglu, Koray. (2015). Spatial Transformer Networks. Advances in Neural Information Processing Systems 28 (NIPS 2015).

[18] S. R. DeGraaf, ”Sidelobe reduction via adaptive FIR filtering in SAR imagery,” in IEEE Transactions on Image Processing, vol. 3, no. 3, pp. 292-301, May 1994, doi: 10.1109/83.287022.