In this article, we propose a novel concept of cross-learning in order to improve synthetic aperture radar (SAR) images by learning from the camera images, in the manifold domain. We present multilevel abstraction approaches to materialize knowledge transfer between these two very different modalities (i.e., the radar and the camera), namely, a canonical correlation analysis-based approach and a manifold alignment-based approach. We provide experimental results on real data, along with qualitative as well as quantitative analyses, to validate the proposed methodologies.

I. INTRODUCTION

Synthetic aperture radar (SAR) [1]–[3] can provide high-resolution images. Substantial amount of work is available to enhance SAR image quality in terms of denoising and/or superresolution (see e.g., [4], [5]). SAR is emerging as a new imaging mode for automotive scenarios [6]–[8]. However, most of the previous work focuses on improving the SAR image by assuming SAR to be a stand-alone sensor, without any interaction with any sensor of a different modality. In automotive (especially, autonomous driving) scenarios, a car may be equipped with multiple sensors, e.g., radar, lidar, camera, etc. [9], [10]. Therefore, it is natural to explore if a SAR image can be improved by using images from other sensors of different modalities, i.e., exploiting the framework of multimodal fusion [11]–[13]. This motivation also forms the basis for our present article.

Multimodal fusion is a very generic concept, combining data/information from diverse modalities in order to enhance the achievement of a common objective, e.g., creating a unified sensing system, improving decision-making,
The following are the main contributions and visualization, fusion, etc. [21]–[25]. However, most of these tasks are (best) carried out in the manifold domain without the aim of reconstructing the high-dimensional data. In our case, we need to, 1) create the respective manifolds of SAR and camera images to generate the intrinsic-dimensional or manifold-domain representation, 2) learn extra features from camera manifold and transfer it to the SAR manifold, 3) reconstruct the SAR image in its high-dimensional or image-domain representation. Since we do the learning in the manifold-domain and then reconstruct the original image-domain, we can only use linear manifolds, e.g., principal component analysis (PCA). Nonlinear manifolds, e.g., Laplacian eigenmaps (LE), locally linear embedding (LLE), Hessian eigenmaps, diffusion maps, etc., provide efficient low-dimensional representation. However, they cannot be transformed/projected back to the image domain. Manifold alignment (MFA) [24], [26], [27] has been an effective way of transferring knowledge/information between different datasets. Similarly, in the case of superresolution of face images, building on a two step approach of global and then local features adjustment [28], [29], some authors, e.g., [30], [31] have advocated the creation of a coherent subspace over the manifolds for efficient transfer of knowledge. Due to an extra layer of abstraction, the latter approach has the ability to work with a modest amount of training samples as well as to compensate for choosing a linear manifold instead of a nonlinear manifold (if required). Now, in case of transferring information from a camera image to a radar image, i.e., cross-learning, there are multiple challenges, e.g., the modalities are different, coordinate systems are different therefore the two sensors cannot be fully registered with each other, there is substantial disparity in resolution, choice of manifolds is limited due to reconstruction requirement, etc. Thus, in order to circumvent these challenges, a multilevel abstraction may be the right course of action. To this end, we follow the approach of [30]. We create PCA-based manifolds for both the sensors and generate a coherent subspace by using the canonical correlation analysis (CCA) [32], [33]. Then, we use LLE [21] to learn/adjust the neighborhood embedding of the coherent subspace from the camera to the radar, followed by recovering the improved SAR image. Following the same logic, we also propose to extend the MFA approach of [27] by incorporating an LLE abstraction layer on the aligned PCA-based manifolds. Thus, rendering the MFA approach as a multilevel abstraction approach. Note, the input SAR images are generated by our recently proposed forward-scanning SAR (FS-SAR) [6] mode for the automotive scenarios, albeit, the synthetic aperture considered here is circular instead of linear, i.e., a circular-scanning SAR (CiS-SAR). Further, in this article, we consider stationary targets, so that the presented work focuses on different aspects of the proposed method. However, future work will include scenarios with moving targets. In this regard, some of our recently published work in [7] may be useful in focusing the targets at a specific position to help in generating stable manifolds.

Contributions: The following are the main contributions of this article.
1) We present a novel concept of improving SAR images via cross-learning from camera images.
2) We show that multiple levels of abstraction can help circumvent the challenges of knowledge transfer in these different modalities.
3) We consider a CiS-SAR mode generated SAR images as input to the cross-learning framework.
4) We present performance results based on real-data obtained in our lab controlled experimental setup.

Advantages/Merits of the Proposed Method. The following are some of the expected advantages or merits of using our proposed method.

1) Cross-learning helps in obtaining high-resolution SAR images which essentially helps in reconstructing the physical details of the extended targets. Thus, the resulting SAR images, having the benefit of both high-resolution as well as detailed information of the target, may possibly replace camera-only images for subsequent processing, e.g., segmentation, identification, or classification.
2) Through the use of manifolds, we are able to provide data level fusion between the SAR and camera images without strict registration, given that the two sensors share the same platform in viewing the same target. Thus, our approach is quite practical.
3) Although we use CiS-SAR images as an example for short-range imaging, other SAR imaging methods can also be used in our cross-learning approach. Thus, our approach has the benefit of flexibility.

Organization: Section II provides the system model, Section III elaborates on the realization of cross-learning via different abstraction approaches, Section IV provides experimental results and performance comparisons, and finally, Section V concludes the article.

Notations: Matrices are in upper case bold while column vectors are in lower case bold, $(\cdot)^T$ denotes transpose, $[a]_i$ is the $i$th element of $a$ and $[A]_{i,j}$ is the $ij$th element of $A$. $\hat{a}$ is the estimate of $a$. $\hat{\Theta}$ defines an entity, and the $\ell_p$-norm is denoted as $||a||_p = \left(\sum_{i=1}^{N} |[a]_i|^p\right)^{1/p}$.

II. SYSTEM MODEL

In [6], we proposed an FS-SAR mode to enhance the azimuth resolution of an automotive radar. This mode combines forward-scanning with SAR processing. FS-SAR assumes a linear aperture. In the present article, we consider a circular aperture, i.e., a circular-scanning SAR (CiS-SAR). CiS-SAR combines the benefits of scanning and spotlight SAR, resulting in enhanced azimuth resolution. We opt for CiS-SAR as a generic SAR mode in order to exhibit the cross-learning possibilities from camera to SAR. However, future works may consider other SAR modes as well. Fig. 1 shows the schematic for CiS-SAR. At each scan-step, $l \in [1, L]$, over the circular aperture, the radar scans the extended target over the angular range (field of view of the sensor), $\theta \in [\theta_{\min}, \theta_{\max}]$. Thus, the target information is obtained both over the circular aperture as well as over the angular scan per aperture position. Similar to compressed sensing-based back-projection (CBP) in [6], we first process the measurements received over the scans by using a compressed sensing-based algorithm to improve the resolution and then back-project the reconstructed images from all the scans over the circular aperture to generate a coherent image of the extended target.

Let, the radar transmit frequency modulated continuous wave (FMCW) pulses toward the target. The signal received is then dechirped, low-pass filtered, deskewed, and Fourier transformed along the fast-time to obtain the range profile (see [6] for explicit expressions and subsequent details on the radar signal model). However, the received signal along the azimuth, for a scanning radar, at scan-step $l$ and range $r$ can be considered as a convolution of the radar antenna beam, $h(\theta)$, and the azimuth reflectivity function, $x_{l,r}(\theta)$, as in [34]

$$y_{l,r}(\theta) = h(\theta) \ast x_{l,r}(\theta) + v_{l,r}$$

where $\ast$ denotes convolution and $v_{l,r}(\theta)$ represents the additive white Gaussian noise (AWGN). Collecting all of the azimuth samples over $\theta$, equation (1) can be written as

$$y_{l,r} = \textbf{GH} x_{l,r} + v_{l,r}$$

where $\textbf{H}$ is a block-Toeplitz convolution matrix and $\textbf{G}$ is a selection matrix to balance the numeric relation between $N_l \times 1$ vector $x_{l,r}$ and $N_\Theta \times 1$ vector $y_{l,r}$. Note, in the context of azimuth-resolution enhancement, $N_\Theta \gg N_\Theta$. Now, concatenating $y_{l,r}$ over all $N_r$ range bins, we can write (2) as

$$y_l = [I_{N_l} \otimes (\textbf{GH})] x_l + v_l \overset{\Delta}{=} A x_l$$

where $y_l = [y_{l,1}^T, y_{l,2}^T, \ldots, y_{l,N_l}^T]^T$ is an $N_\Theta N_r \times 1$ vector. Similarly, $x_l$ and $v_l$ can be defined as $N_\Theta N_r \times 1$ and $N_\Theta N_r \times 1$ vectors, respectively. Further, $\textbf{A}$ is the $N_\Theta N_r \times N_\Theta N_r$ measurement matrix. Now, according to CBP, $x_l$ (assuming it to be sparse) can be estimated by solving the following (fused LASSO [35]) optimization problem.

$$\hat{x}_l = \arg \min_{x_l} \|y_l - \textbf{Ax}_l\|_2^2 + \lambda_x \|x_l\|_1 + \lambda_f \|D x_l\|_1$$

where $\lambda_x$ and $\lambda_f$ are positive penalty parameters controlling element-wise sparsity and fusion in $x_l$, respectively, and $\textbf{D}$

\[\text{Fig. 1. SAR and camera system schematic.}\]
is the fusion matrix (i.e., $\mathbf{Dx}$) is a vector of differences of consecutive elements of $\mathbf{x}$) \cite{35}. Then, the reconstructed radar image via back-projection, at pixel $(i, j)$, can mathematically be represented as

$$
g_{i,j} = \sum_{l=1}^{L} \mathbb{1}_{i,i'}(\tilde{\mathbf{X}}_l)_{l,b_i,b_j}$$

(5)

where $\mathbb{1}_{i,i'}(\cdot)$ interpolates/upsamples a matrix by an order $\kappa$ and $\kappa'$ along its rows and columns, respectively, $\tilde{\mathbf{X}}_l$ is the $N_x \times N_y$ reshaped matrix form of $\mathbf{x}_l$, $[l,b_i,b_j]$ represents the row and column indices of the matrix corresponding to angle $\theta_{i,j}$ and range $r_{i,j}$ for the $(i, j)$th pixel, respectively. Since each scanning position over the aperture may contribute to each pixel in the reconstructed image, we can rewrite (5) as

$$
g_{i,j} = \sum_{l=1}^{L} \gamma_{i,j}^l$$

(6)

where $\gamma_{i,j}^l \triangleq [\mathbb{1}_{1,i'}(\tilde{\mathbf{X}}_l)]_{l,b_i,b_j}$. Let, all of the image pixels $\gamma_{i,j}^l$, for $i = 1, \ldots, \sqrt{N}$ and $j = 1, \ldots, \sqrt{N}$, w.r.t. contributions from the $l$th aperture position, are collected in an $N \times 1$ vector $\mathbf{r}_l$

$$
\mathbf{r}_l \triangleq \left[\gamma_{1,1}^l, \ldots, \gamma_{1,\sqrt{N}}^l, \gamma_{\sqrt{N},1}^l, \ldots, \gamma_{\sqrt{N},\sqrt{N}}^l\right]^T.
$$

(7)

Now, we can collect all of the radar images generated from each aperture position, as defined in (7), as an $N \times L$ matrix $\mathbf{R}$

$$
\mathbf{R} \triangleq [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_L].
$$

(8)

For the proposed cross-learning, we assume that the camera trajectory is the same as that of the radar synthetic aperture, i.e., the camera images of the target are also obtained from the same physical location as that of the radar. However, the camera does not involve any scanning and takes one snapshot for each location over its trajectory. Fig. 1 shows the schematic of camera image acquisition. For the sake of clarity, the schematic for camera has been drawn separately from SAR. However, in practice, both sensors may share the physical location. The camera may be mono or stereo, with different image formats, i.e. RGB, grayscale, depth-map, etc. It may also have its own requirements w.r.t. configuration, calibration, disparity/cloud-production formulation, etc. We assume that these prerequisites have already been met. However, we do not assume any strict registration between the camera and the radar, as it is very difficult and our approach essentially tries to circumvent its need.

Let, a $\sqrt{M} \times \sqrt{M}$ generic camera image of the target at $l$th position on its trajectory is represented as an $M \times 1$ vector $\mathbf{s}_l$ via lexicographic ordering (column ordered). Then, we can collect all such images for the complete trajectory into an $M \times L$ matrix $\mathbf{S}$ as

$$
\mathbf{S} \triangleq [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L].
$$

(9)

The rest of the article essentially deals with (8) and (9) in terms of proposing a cross-learning strategy.

III. CROSS-LEARNING

In order to improve SAR images via cross-learning from camera images, we present two multilevel abstraction approaches, i.e., 1) a multilevel CCA (ML-CCA) based cross-learning and 2) a multilevel MFA (ML-MFA) based cross-learning. For both the approaches, we first learn the respective manifolds of the training images of the two sensing modalities.

A. Suitable Manifold

As explained earlier, after cross-learning, we need to reconstruct the SAR image from low-dimensional space of the manifold domain to the high-dimensional space of the image domain. Therefore, nonlinear manifolds cannot be used. In terms of linear manifolds, we opt for the classical PCA-based manifolds.

PCA represents data by using the directions of maximum variance. Thus, it requires computing the principal eigenvectors of the data covariance matrix. Now, assuming that the datasets in (8) and (9) are centred (i.e., the corresponding sample means have been subtracted from them), the covariance matrices of radar and camera datasets can be defined as, $\mathbf{C}_r \triangleq (1/L)\mathbf{R}\mathbf{R}^T$ and $\mathbf{C}_s \triangleq (1/L)\mathbf{S}\mathbf{S}^T$, respectively. The eigenvalue decomposition (EVD) of the covariance matrices can then be carried out as

$$
\text{EVD}(\mathbf{C}_r) = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T
$$

(10)

$$
\text{EVD}(\mathbf{C}_s) = \mathbf{U}_s \mathbf{\Sigma}_s \mathbf{V}_r^T
$$

(11)

where matrices $\mathbf{U}_r$ and $\mathbf{U}_s$ contain the left eigenvectors, matrices $\mathbf{V}_r$ and $\mathbf{V}_s$ contain the right eigenvectors, and matrices $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_s$ contain the corresponding eigenvalues along their diagonals, for radar and camera, respectively. The low-dimensional data representation essentially corresponds to projecting the data on a few significant eigenvectors. Let, $n$ and $m$ represent the number of significant eigenvectors (or subsequent principal components) for SAR and camera manifolds, respectively. Then, $\hat{\mathbf{U}}_r \triangleq [\mathbf{U}_r]_{1:n}$ and $\hat{\mathbf{U}}_s \triangleq [\mathbf{U}_s]_{1:m}$ are the $N \times n$ and $M \times m$ corresponding PCA-based projection matrices. The PCA-based projection coefficients can be obtained as

$$
\mathbf{P}_r = \hat{\mathbf{U}}_r^T \mathbf{R} = \begin{bmatrix} \mathbf{p}_{r_1}, \mathbf{p}_{r_2}, \ldots, \mathbf{p}_{r_L} \end{bmatrix}
$$

(12)

$$
\mathbf{P}_s = \hat{\mathbf{U}}_s^T \mathbf{S} = \begin{bmatrix} \mathbf{p}_{s_1}, \mathbf{p}_{s_2}, \ldots, \mathbf{p}_{s_L} \end{bmatrix}
$$

(13)

where $\mathbf{p}_{r_i} \triangleq \hat{\mathbf{U}}_r^T \mathbf{r}_i$, $\mathbf{p}_{s_i} \triangleq \hat{\mathbf{U}}_s^T \mathbf{s}_i$, and, $\mathbf{P}_r$ and $\mathbf{P}_s$ are $n \times L$ and $m \times L$ PCA coefficient matrices of SAR and camera manifolds, respectively.

B. ML-CCA-Based Cross-Learning

In ML-CCA-based cross-learning, we first use CCA to generate a coherent subspace between the two modalities. Then, we use LLE for neighborhood embedding w.r.t. test images of the radar and camera. Finally, the processed radar image is projected from the manifold domain back to the image domain.

1) Coherent Subspace: CCA finds a low-dimensional coherent subspace between two datasets. In this article, we
consider a 1-D CCA subspace. Thus, in our case, CCA provides one basis vector for each dataset, such that the correlation between the corresponding projection coefficients is maximized. Note, the datasets, in our case, correspond to PCA-based manifold coefficients, i.e., equations (12) and (13). Mathematically, we can estimate the CCA-based subspace by solving the following optimization problem, as in [32]

\[
\begin{bmatrix} b_r, b_s \end{bmatrix} = \arg \max_{b_r, b_s} b_r^T Q_r b_r b_s^T Q_s b_s = 1 \quad \text{s.t.} \quad b_r^T Q_r b_r = 1, \quad b_s^T Q_s b_s = 1
\]

(14a)

(14b)

where \( b_r \) and \( b_s \) are \( n \times 1 \) and \( m \times 1 \) canonical basis vectors for the radar- and camera-manifold datasets, respectively, and \( Q_r = (1/L)P_rP_r^T \), \( Q_s = (1/L)P_sP_s^T \) and \( Q_{rs} = (1/L)P_rP_s^T \) are the corresponding covariance matrices. Note, the constraints (14b) are imposed to ensure a unique solution. Now, solving (14) essentially boils down to solving the following generalized eigenvalue problem (see [32] for details)

\[
\begin{bmatrix} Q_r & 0 \\ 0 & Q_s \end{bmatrix} \begin{bmatrix} b_r \\ b_s \end{bmatrix} = \lambda \begin{bmatrix} 0 & Q_s \\ Q_r & 0 \end{bmatrix} \begin{bmatrix} b_r \\ b_s \end{bmatrix}
\]

(15)

where \( \lambda \) is the generalized eigenvalue. Solving (14) is equivalent to finding the largest generalized eigenvalue in (15), i.e., \( \lambda = \lambda_{\text{max}} \), and the corresponding generalized eigenvector provides the estimate of canonical basis vectors as, \( \begin{bmatrix} b_r^T, b_s^T \end{bmatrix}^T \). From the basis vectors, the corresponding CCA-based coefficients can be obtained as

\[
\begin{align*}
\mathbf{a}_r &= P_r^T \hat{b}_r \\
\mathbf{a}_s &= P_s^T \hat{b}_s
\end{align*}
\]

(16)

(17)

where \( \mathbf{a}_r \) and \( \mathbf{a}_s \) are \( L \times 1 \) vectors of CCA-based coefficients w.r.t. the radar and camera manifolds, respectively.

2) Neighborhood Embedding: LLE is used to compute low-dimensional neighborhood-preserving embeddings of high-dimensional data. It is based on a simple geometric intuition. Given a data point and its neighbors, in high-dimension, lie on a locally linear patch of the manifold, the data point can be reconstructed by linear combination of its neighbors. Then, the data point can be mapped to a low-dimensional representation while preserving its neighborhood characterization (see [21] for more details). In the context of cross-learning, 1) the mapping is done from the camera manifold to the radar manifold, 2) the radar and the camera manifolds have been substituted with respective CCA-based coefficients which are linear due to one-dimensional CCA subspace, and 3) in terms of CCA-based coefficients, the data dimension for both radar and camera is the same, therefore, we do not need to find the low-dimensional values. Thus, LLE can be easily applied to our case for neighborhood embedding, i.e., we need to find the linear coefficients which reconstruct a camera data point from its neighbors and then use the same linear coefficients to reconstruct a radar data point from its neighbors. This constitutes neighborhood embedding in the context of cross-learning.

Let, \( \mathbf{r}_i \) and \( \mathbf{s}_i \) be the test SAR and camera images, respectively, with \( \mathbf{p}_r = \hat{U}_r \mathbf{r}_i \) and \( \mathbf{p}_s = \hat{U}_s \mathbf{s}_i \) as the corresponding data points on the manifolds. Then, the CCA-based coefficients for the test images can be obtained as

\[
\begin{align*}
\mathbf{a}_{r_i} &= \mathbf{p}_r^T \hat{b}_r \\
\mathbf{a}_{s_i} &= \mathbf{p}_s^T \hat{b}_s
\end{align*}
\]

(18)

(19)

where \( \mathbf{a}_{r_i} \) and \( \mathbf{a}_{s_i} \) are scalar values. Let, \( \mathcal{N}_r^k \) and \( \mathcal{N}_s^k \) represent the sets of \( K \) nearest neighbors (K-NN) of \( \mathbf{a}_{r_i} \) and \( \mathbf{a}_{s_i} \), respectively. Now, we can write the optimization problem of finding the linear coefficients of reconstructing \( \mathbf{a}_{r_i} \) from its neighbors in \( \mathbf{a}_s \) as a constrained least-squares fitting problem

\[
\hat{\mathbf{w}} = \arg \min_w \| \mathbf{a}_{r_i} - \mathbf{w}^T \bar{\mathbf{a}}_s \|_2^2 \quad \text{s.t.} \quad \| \mathbf{w} \|_2^2 = 1
\]

(20a)

(20b)

where \( \bar{\mathbf{a}}_s \) is \( K \times 1 \) subvector of \( \mathbf{a}_s \), such that \( \mathbf{a}_{s_i} \in \mathcal{N}_s^k \), for \( i = 1, \ldots, K \). Note, equation (20) essentially applies two constraints, 1) a sparseness constraint, i.e., weights are nonzero only for the K-NN of \( \mathbf{a}_{r_i} \), 2) an invariance constraint, i.e., the sum of linear coefficients equals one, as (20b). An efficient way to minimize the error in (20a) is to solve the following system of linear equations:

\[
\mathbf{G} \hat{\mathbf{w}} = \mathbf{1}
\]

(21)

where \( \mathbf{G} \triangleq (\mathbf{a}_{r_i} - \bar{\mathbf{a}}_s)(\mathbf{a}_{r_i} - \bar{\mathbf{a}}_s)^T \), and then rescale the coefficients to satisfy (20b) (more details in [21]). Now, the learnt coefficients can be used to reconstruct the radar data point from the neighbors

\[
\hat{\mathbf{a}}_{r_i} = \hat{\mathbf{w}}^T \bar{\mathbf{a}}_s
\]

(22)

where \( \bar{\mathbf{a}}_s \) is \( K \times 1 \) subvector of \( \mathbf{a}_s \), such that \( \mathbf{a}_{s_i} \in \mathcal{N}_s^k \), for \( i = 1, \ldots, K \).

3) Image Reconstruction: After learning the CCA-based coefficient, the learnt radar image in the manifold domain can be obtained as

\[
\hat{\mathbf{r}}_i = (\hat{\mathbf{b}}_r^T)^T \hat{\mathbf{a}}_{r_i} + \mathbf{r}_i
\]

(23)

where \( (\cdot)^T \) denotes the Moore–Penrose (or pseudo) inverse. Now, the radar image can be projected from the manifold domain back to the image domain as

\[
\mathbf{f}_i = \hat{U}_r \hat{\mathbf{r}}_i
\]

(24)

where \( \hat{f}_i \) is the improved SAR image obtained via cross-learning from the camera image.

C. ML-MFA-Based Cross-Learning

In this section, we present ML-MFA-based cross-learning. First, we reconstruct an image after MFA only, and then, we introduce the abstraction layer of neighborhood embedding via LLE to distinguish between low-level abstraction and multilevel abstraction.

1) MFA: In this approach the PCA-based manifolds of the two sensors are essentially aligned using Procrustes analysis as in [27]. The basic idea is that given pairwise correspondence between the two datasets (assumed centred), a
mapping is obtained to align the test data points. Using the earlier terminology developed in this article, the following singular value decomposition (SVD) is performed as a first step

\[
\text{SVD}(\mathbf{P}_r \mathbf{P}_r^T) = \mathbf{U} \Sigma \mathbf{V}^T
\]

(25)

assuming \( n = m \). Then, MFA is obtained as

\[
\hat{\mathbf{p}}_r = k \mathbf{Q}^T \mathbf{p}_r
\]

(26)

where \( k \) \( = \text{trace}(\Sigma)/\text{trace}(\mathbf{P}_r \mathbf{P}_r^T) \) and \( \mathbf{Q} = \mathbf{U} \mathbf{V}^T \), and the MFA-based SAR image can be obtained via (24) by using the relation, \( \hat{\mathbf{p}}_r = \hat{\mathbf{p}}_r \).

2) Neighborhood Embedding: Now, we extend the MFA by neighborhood embedding via LLE (similar to Section III-B2). This essentially forms the ML-MFA. Mathematically, it can be represented by rewriting (18) and (19) as

\[
\mathbf{a}_r = \hat{\mathbf{p}}_r, \quad \mathbf{a}_s = \mathbf{p}_s
\]

(27)

where \( \mathbf{a}_r \) and \( \mathbf{a}_s \) are vectors, and solving

\[
\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \| \mathbf{a}_s - \tilde{\mathbf{A}}_s \mathbf{w} \|_2^2 \text{ s.t. } \| \mathbf{w} \|_2 = 1
\]

(28)

where \( \tilde{\mathbf{A}}_s \) is a matrix containing \( K \) neighbors of \( \mathbf{a}_s \) in \( \mathbf{P}_s \), as its columns, and rewriting (22) as

\[
\hat{\mathbf{a}}_r = \tilde{\mathbf{A}}_r \hat{\mathbf{w}}
\]

(29)

where \( \tilde{\mathbf{A}}_r \) is a matrix containing \( K \) neighbors of \( \mathbf{a}_r \) in \( k \mathbf{Q}^T \mathbf{P}_r \), as its columns. Then, by using the relation \( \tilde{\mathbf{p}}_r = \hat{\mathbf{a}}_r \), SAR image can be obtained via (24).

IV. EXPERIMENTAL RESULTS

In this section, we validate the proposed concept of cross-learning between SAR and camera images by experimental results. According to our knowledge, a public dataset for concurrent SAR and camera measurements is not available. Therefore, as part of this research, we have carried out such measurements in a laboratory controlled environment.\(^1\) Our experimental setup (see Fig. 2) mainly consists of a ZED stereo camera (see [36] for specifications), a 300-GHz FMCW radar (see Table I for specifications) and a trolley (of size \( 1 \times 0.5 \times 0.55 \) (length \( \times \) width \( \times \) height) m\(^3\) ) on a turntable. Fig. 3 shows the measurement schematic of the experiment. The trolley is placed on a turntable at a distance of 3.65 m from the joint sensors (ZED and radar) platform. Measurements from the sensors are taken for every 5° angular turn (counter-clock-wise) of the trolley. This emulates the circular motion of the sensors around the target. Thus, \( L = 72 \) (synthetic) aperture samples are obtained all around the trolley. Note, in this article, we consider a full 360° circular aperture for the purpose of illustration only. However, in practice, the target can be seen by a partial aperture. For every aperture sample, we consider only left lens RGB image from the ZED. We convert the RGB image to a grayscale image. Fig. 4 shows the ZED images of the trolley at different positions (on the turntable) \( l = 1 \) (0°), 19 (90°), 37 (180°), 55 (270°). At every aperture position, the radar scans the target scene for an angular range \( \theta = \pm 13^\circ \), at angular intervals \( \Delta \theta = 0.25^\circ \). Then, using (4) and (7), a SAR image is created for every \( l \) th aperture position. Finally, a combined image of the target is achieved by using (5). Fig. 5 shows the SAR images for \( l = 1 \) (0°), 19 (90°), 37 (180°), 55 (270°) and the combined CiS-SAR image of the trolley. Note, all SAR images have

\[^1\]All real-data supporting this work are openly available from The University of Edinburgh repository (DataShare) at https://doi.org/10.7488/ds/2828

| TABLE I Specifications of 300 GHz Radar |
|----------------------------------------|
| **Modulation** | **FMCW** |
| **Frequency Range** | 287 – 293 GHz |
| **Transmit Bandwidth (B)** | 5 GHz |
| **Chirp Duration (T)** | 1 ms |
| **Angular Step (\( \Delta \theta \))** | 4.096 MHz |
| **Range Resolution (\( \Delta r \))** | 0.25° |
| **Two-way 3 dB Beamwidth (\( \theta_{3db} \))** | 13° |

![Fig. 3. Measurement schematic.](image)
been normalized so that the maximum intensity is unity. We can see that the SAR images of individual apertures, Fig. 5(a)–(d), capture viewing-angle dependent information of the target. Nonetheless, the combined image, Fig. 5(e), provides very good imaging result in capturing the complete outline of the target, which reaffirms the enhanced performance of CBP reconstruction algorithm as proposed in [6]. However, we can see that the handle of the trolley is not very prominent.

Now, in order to improve SAR imaging results we use the cross-learning concept as explained in Section III. Using (12) and (13), for \( n = m = 15 \) principal components, we obtain the PCA-based manifolds of the images of the two sensors. Note, we chose these number of principal components as they seemed to provide good results from a qualitative perspective. However, given the training data, a better estimate can easily be found. Nonetheless, in our experience, changing these numbers does not have a drastic impact on the trend of performance, as would become clear shortly. Fig. 6 shows these manifolds (as scatter plots) for first two principal components. We can see that the ZED manifold is more elaborate than the SAR manifold (which is quite concentrated). This shows that the ZED images are more distinguishable than the SAR images. Thus, the SAR images have a big margin of learning from the camera images.

Fig. 7(a) shows the SAR image using ML-CCA approach. Note, we essentially use all of the training images as the test images, i.e., \( t = 1, \ldots, L \), for both the sensors. We can see that in comparison to CiS-SAR [Fig. 5(e)], ML-CCA image has captured some new information of the target. The walls of the trolley are more prominent. However, the most interesting aspect is the visibility of the trolley handle. Nonetheless, we can see that the target information both in CiS-SAR and ML-CCA does not seem to overlap for every pixel. Thus, a natural course of action is to combine the two images. We name the combined CiS-SAR and ML-CCA image as ML-CCA+. Fig. 7(b)
shows the ML-CCA+ image. We can see that it is a much improved image than the CiS-SAR as in Fig. 5(e).

Fig. 7(c) shows the imaging result of MFA approach (which uses low-level of abstraction). Similar to ML-CCA+, we also provide the image result of MFA+ in Fig. 7(d). We can see that MFA does get some extra information in comparison to CiS-SAR [Fig. 5(e)]. However, its performance is inferior to ML-CCA [Fig. 7(a)]. Similarly, ML-CCA+ [Fig. 7(b)] provides better result than MFA+.

Fig. 7(e) shows the imaging result of ML-MFA. Similar to ML-CCA+, we also provide the image result of ML-MFA+ in Fig. 7(f). We can see that ML-MFA is also showing some features of the trolley handle. Therefore, ML-MFA+ shows an improved image. It is better than MFA+ [Fig. 7(d)], since it uses multilevel abstraction.

Fig. 8 provides images, zoomed in on the trolley front-end, created by CiS-SAR, MFA+, ML-MFA+, and ML-CCA+, to better assess the qualitative comparison. Note, we have also used Matlab’s smoothing function to reduce pixelation (for better display). We can see that cross-learning has helped in collecting more information of the target. In particular, ML-MFA+ and ML-CCA+ which use multilevel abstraction, show more reflectivity and structure of the trolley (i.e., walls and handle) than CiS-SAR.

Now, in order to provide quantitative analyses of the presented methods, we need a ground truth which is not easy to obtain. To circumvent this, we consider a pseudo ground-truth. We estimate the noise level in the generated image of each method via a histogram (i.e., the most populated bin in a histogram of 50 bins) and remove it from the respective image. Then, we quantify all reflectivity values to unity. This provides us a pseudo ground-truth, albeit favoring its generative method. Fig. 9 shows the resulting images. We can see that with small variations, all the images look quite similar. Thus, these can serve as approximate ground truths for all the methods. We basically take each of these images and then compare the performance of all methods with it. For quantitative comparison, we use Matlab’s function $ssim$ which is a structural similarity index metric for image comparison, i.e., it combines local image structure, luminance, and contrast into a single local quality score. In our case, it would basically measure the structure and total reflectivity of each image in comparison to the ground truth. Note, 1 is the highest $ssim$ score. Table II provides the comparison results. The columns represent the ground truth image used and the rows provide the similarity measure for each method. We can see that for all the cases, multilevel abstraction-based methods, i.e., ML-MFA+ and ML-CCA+, perform much better than the others. MFA+ basically uses low-level of abstraction, but still outperforms CiS-SAR. Table II provides the comparison results for the PCA order $n = m = 15$. We also present results by varying PCA order to $n = m = 5$ and $n = m = 25$, in order to see the impact of such variations, in Tables III and IV, respectively.
We can see that there is not much difference in the general trend of performance. However, given the training data, an appropriate value of PCA order can always be found.

From above, we can say that multilevel abstraction approaches have superior performance in realizing the concept of cross-learning to improve SAR images by using the camera images.

V. CONCLUSION

In this article, we have proposed a novel concept of cross-learning, in order to improve SAR images by learning from the camera images. Despite the fact that the two sensors are very different modalities, we have used multilevel abstraction approaches to achieve knowledge transfer between them. To this end, we have presented a CCA- and a MFA-based multilevel abstraction approach. In order to validate the proposed concept, we have provided experimental results on real data. Through qualitative and quantitative analyses, we have shown that multilevel abstraction approaches provide better performance results in terms of collecting more information about the target structure.

REFERENCES

[1] W. Carrara, R. Goodman, and R. Majewski
Spotlight Synthetic Aperture Radar. Boston, MA, USA: Artech House, 1995.

[2] C. Jakowitz, D. Wahl, P. Eichel, D. Ghiglia, and P. Thompson
Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach. Norwell, MA, USA: Kulwer, 1996.

[3] M. Soumekh
Synthetic Aperture Radar Signal Processing With MATLAB Algorithms. New York, NY, USA: Wiley, 1999.

[4] M. Cetin et al.
Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocusing, moving targets, and compressed sensing
IEEE Signal Process. Mag., vol. 31, no. 4, pp. 27–40, Jul. 2014.

[5] S. Gishkori and B. Mulgrew
Graph signal processing-based imaging for synthetic aperture radar
IEEE Geosci. Remote Sens. Lett., vol. 17, no. 2, pp. 232–236, Feb. 2020.

[6] S. Gishkori, L. Daniel, M. Gashinova, and B. Mulgrew
Imaging for a forward scanning automotive synthetic aperture radar
IEEE Trans. Aerosp. Electron. Syst., vol. 55, no. 3, pp. 1420–1434, Jun. 2019.

[7] S. Gishkori, D. Wright, L. Daniel, M. Gashinova, and B. Mulgrew
Imaging moving targets for a forward-scanning automotive SAR
IEEE Trans. Aerosp. Electron. Syst., vol. 56, no. 2, pp. 1106–1119, Apr. 2020.

[8] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian
The rise of radar for autonomous vehicles: Signal processing solutions and future research directions
IEEE Signal Process. Mag., vol. 36, no. 5, pp. 20–31, Sep. 2019.

[9] E. Guizzo
How google’s self-driving car works
Oct. 2011. [Online]. Available: https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works

[10] H. Cho, Y. Seo, B. V. K. V. Kumar, and R. R. Rajkumar
A multi-sensor fusion system for moving object detection and tracking in urban driving environments
In Proc. IEEE Int. Conf. Robot. Automat., May 2014, pp. 1836–1843.

[11] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi
Multisensor data fusion: A review of the state-of-the-art
Inf. Fusion, vol. 14, no. 1, pp. 28–44, Jan. 2013.

[12] D. Lahat, T. Adali, and C. Jutten
Multimodal data fusion: An overview of methods, challenges, and prospects
Proc. IEEE, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[13] D. Ramachandram and G. W. Taylor
Deep multimodal learning: A survey on recent advances and trends
IEEE Signal Process. Mag., vol. 34, no. 6, pp. 96–108, Nov. 2017.

[14] L. Sorber, M. Van Barel, and L. De Lathauwer
Structured data fusion
IEEE J. Sel. Topics Signal Process., vol. 9, no. 4, pp. 586–600, Jun. 2015.

[15] C. Debes et al.
Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest
IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2405–2418, Jun. 2014.

[16] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson
Challenges and opportunities of multimodality and data fusion in remote sensing
Proc. IEEE, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[17] S. J. Pan and Q. Yang
A survey on transfer learning
IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[18] C. M. Bishop
Pattern Recognition and Machine Learning (Information Science and Statistics). Berlin, Heidelberg: Springer-Verlag, 2006.

[19] X. Huo and A. Smith
A Survey of manifold-based learning methods
In Recent Advances in Data Mining of Enterprise Data, Singapore: World Scientific, pp. 691–745, 2008.

[20] L. van der Maaten, E. Postma, and J. van den Herik
Dimensionality reduction: A comparative review
Tilburg Univ., Tilburg, the Netherlands, Tech. Rep. TiCC TR 2009-005, Oct. 2009.

[21] S. T. Roweis and L. K. Saul
Nonlinear dimensionality reduction by locally linear embedding
Science, vol. 290, pp. 2323–2326, 2000.

[22] M. Belkin and P. Niyogi
Laplacian eigenmaps for dimensionality reduction and data representation
Neural Comput., vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[23] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk
The multiscale structure of non-differentiable image manifolds
In Proc. SPIE Conf. Series, 2005, pp. 413–429.
[24] S. Lafon, Y. Keller, and R. R. Coifman
Data fusion and multicue data matching by diffusion maps
IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 11,
pp. 1784–1797, Nov. 2006.

[25] M. A. Davenport, C. Hegde, M. F. Duarte, and R. G. Baraniuk
Joint manifolds for data fusion
IEEE Trans. Image Process., vol. 19, no. 10,
pp. 2580–2594, Oct. 2010.

[26] J. Ham, D. Lee, and S. Lawrence
Semisupervised alignment of manifolds
In Proc. Ann. Conf. Uncertainty Artif. Intell., vol. 10,
Jan. 2005, pp. 120–127.

[27] C. Wang and S. Mahadevan
Manifold alignment using procrustes analysis
In Proc. 25th Int. Conf. Mach. Learn., 2008,
pp. 1120–1127.

[28] C. Liu, H.-Y. Shum, and W. T. Freeman
Face hallucination: Theory and practice
Int. J. Comput. Vis., vol. 75, no. 1,
pp. 115–134, Oct. 2007.

[29] H. Chang, Dit-Yan Yeung, and Y. Xiong
Super-resolution through neighbor embedding
In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.,
vol. 1, Jun. 2004, pp. 1–8.

[30] H. Huang, H. He, X. Fan, and J. Zhang
Super-resolution of human face image using canonical correlation analysis
Pattern Recognit., vol. 43, no. 7,
pp. 2532–2543, Jul. 2010.

[31] H. Huang and H. He
Super-resolution method for face recognition using nonlinear mappings on coherent features
IEEE Trans. Neural Netw., vol. 22, no. 1,
pp. 121–130, Jan. 2011.

[32] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor
Canonical correlation analysis: An overview with application to learning methods
Neural Computat., vol. 16, no. 12,
pp. 2639–2664, Dec. 2004.

[33] A. J. Izenman
Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning,
1st ed. Berlin, Germany: Springer, 2008.

[34] M. A. Richards
Fundamentals of radar signal processing. New York, NY, USA:
McGraw-Hill, 2005.

[35] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight
Sparsity and smoothness via the fused LASSO
J. Roy. Stat. Soc. Ser. B, vol. 67, no. 1,
pp. 91–108, 2005.

[36] Stereolabs, “ZED stereo Camera,” 2019. [Online]. Available:
https://www.stereolabs.com/