IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 5, 2019, accepted April 25, 2019, date of publication May 1, 2019, date of current version May 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914273

Robust Image Translation and Completion Based on Dual Auto-Encoder With Bidirectional Latent Space Regression

SUKHAN LEE, (Fellow, IEEE), AND NAEEM UL ISLAM

Department of Electrical and Computer Engineering, Intelligent Systems Research Institute, Sungkyunkwan University, Suwon 440-746, South Korea

Corresponding author: Sukhan Lee (lsh1@skku.edu)

This work was supported in part by the 3D Recognition Project of Korea Evaluation Institute of Industrial Technology (KEIT) under Grant 10060160, in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean Government(MSIT) by AI Graduate School Support Program under Grant 2019-0-00421, and in part by the Robot Industry Fusion Core Technology Development Project of KEIT, sponsored by the Korea Ministry of Trade, Industry and Energy (MOTIE) under Grant 10048320.

ABSTRACT Automated image translation and completion is a subject of keen interest due to their impact on image representation, interpretation, and enhancement. To date, a conditional or a dual adversarial framework with a convolutional auto-encoder embedded as a generator is known to offer the best accuracy in image translation. However, although the frequency is excellent, the adversarial framework may suffer from a lack of generality, i.e., the accuracy drops when translating incomplete and corrupted data given as untrained noisy input data. This paper proposes an approach to robust image-to-image translation that offers a high level of generality while keeping accuracy high as well. The proposed approach is referred to here as a dual auto-encoder with bidirectional latent space regression or Bi-directionally Associative DualAE, for short. The proposed BA-DualAE is configured with two auto-encoders the individual latent spaces of which are tightly associated by a bidirectional regression network. Once the two auto-encoders are trained independently for their respective domains, and then, the bidirectional regression network is trained to learn the general association between data pairs. With its capability of robust and bidirectional image translation, BA-DualAE performed direct image completion with no iterative search. The experiments with photo-sketch datasets demonstrated that the proposed BA-DualAE is highly robust under incomplete or corrupted data conditions and is far superior to adversarial frameworks in terms of generality and robustness.

INDEX TERMS Convolutional neural network, auto-encoder, bidirectional latent space regression, image-to-image translation, generative adversarial network.

I. INTRODUCTION

Automated translation of images has implications for a broad range of image processing tasks, including image coloring, styling, de-noising, modification, and completion. In particular, the deep learning approach to image translation is to learn a general association embedded in pairwise data that can affect data interpretation and enhancement. Recent advances in deep learning networks, for instance, conditional and dual adversarial frameworks with convolutional auto-encoders embedded as generators, together with the availability of a wide range of training data sets, offer powerful means of achieving high-quality image-to-image translation. This is because the aim of an adversarial framework is to force the convolutional auto-encoder to generate its outputs as close as possible to the distribution specified by the training data set. However, although these generative adversarial network (GAN) based frameworks are high in accuracy, there remains an important issue to be addressed: achieving sufficient generality while keeping accuracy high or, in short, achieving high robustness in image translation. For instance, our experiments show that the image translation based on a conditional GAN (cGAN) framework tends to carry a corrupted or occluded part of the input image to the corresponding output image in translation unless the network is trained explicitly with such corrupted or occluded images a priori. To achieve high robustness in image translation and completion, this paper proposes a framework of dual auto-encoders the latent space of which are bi-directionally connected by an association network. The proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Shorif Uddin.

framework exploits the capability of the latent space association network for generalization, while taking advantage of dual auto-encoders for the accurate reconstruction of input images independently in their own domains. For more details, refer to the Section III: Proposed Approach.

The rest of the paper is organized as follows: Section II and Section III present, respectively, the related work and the proposed Bi-directionally Associative DualAE framework for cross-domain image-to-image translation in more detail. Section IV and Section V explain the details of the proposed network architecture and the training procedure of the proposed Bi-directionally Associative DualAE framework for image-to-image translation and completion. Finally, Section VI presents the comparative performance analysis by experiments for corrupted and occluded testing samples, before the conclusion in Section VII.

II. RELATED WORK

Within-domain image translation has applications in domain adaptation [1]–[6], super-resolution [7], style transfer [8], and photo editing [9], Target and anomaly detection [10]–[13], and cross-domain image translation has applications in data generation [14], data interpretation [15], transformation of 3D images to their corresponding 3D representation for interpretation of deep CNN [16], and image completion [15], [17], [18]. The availability of a large amount of paired data for image translation makes convolutional neural network (CNN) approaches to regression highly attractive for both within- and cross-domain image translation, surpassing the performance of the state-of-the-art non-CNN approaches [19], [20]. A number of deep generative networks, such as autoencoder (AE) [21], variational auto-encoders (VAEs) [22], [23], generative adversarial network [14], moment matching networks [24], pixel-CNN [25], and plug-and-play generative networks [26], have been proposed that are to learn the distribution of input data for generating realistic images. Recently, many variants of deep auto-encoders as well as of GAN have been proposed, including [22], [23], LapGAN [27], DCGAN [28], WGAN [29] and conditional generative adversarial network (cGAN) [15]. However, it is the combination of AE and GAN that has shown the best performance in automated image translation [15]. Shrivastava *et al.* [17] addressed the problem of performance degradation to real images after training on synthetic images and proposed translating synthetic images to real images using a conditional generative adversarial network, or cGAN, in which the L_1 distance between synthetic and real images is minimized together with the adversarial loss. Isola *et al.* [15] proposed training a pix2pix cGAN to represent the exact pixel correspondence of images for translation.

In contrast, Yi *et al.* proposed dualGAN [18], where a dual configuration of GAN allows the network to be not only bidirectional in image translation but also trainable in an unsupervised way without the need of explicit correspondences.

To assess various deep learning-based approaches to image translation, we define the following criteria: 1) accuracy measuring the translation error from the ground truth, 2) generality representing the dependency of accuracy on untrained noisy data such as incomplete and corrupted data, and 3) utility indicating the ease of training and the flexibility in application. First, the conditional adversarial framework with a convolutional AE or its variant used as its generator has been shown to be superior to direct translation based on GAN or AE alone [21], [28]. Notably, the dual configuration of GAN with forward and inverse paths has been reported to show the highest accuracy to date, especially for cross-domain translation [15].

However, conditional and dual adversarial frameworks as well as direct translation based on AE or GAN have shown limitations in generality [15]; that is, they suffer from a sudden loss of accuracy under conditions of incomplete or corrupted data unless they are trained explicitly [30]. These frameworks tend to weigh more on accuracy than on generality such that the incomplete or corrupted part of the input is undesirably incorporated into the corresponding output. To reduce the burden of preparing for a large scale of noisy or labeled data for training, approaches that provide high robustness or allow unsupervised learning such as dualGAN [18] are preferred. In addition, approaches capable of bidirectional image translation [18] can extend their applicability to more diverse translation problems.

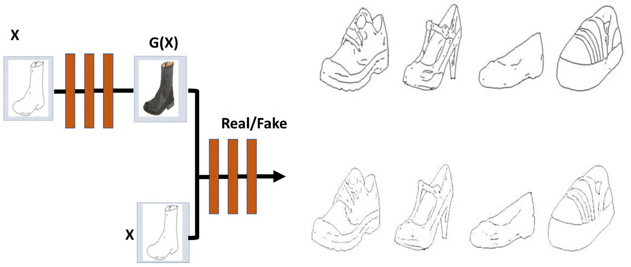


FIGURE 1. cGAN based Image-to-Image translation [15]. Left: A schematic of cGAN implemented. Right: Ground truth images (Top) and cGAN generated images (Bottom).

A. PROBLEM SETTING

We implemented a well-known cGAN approach to image-to-image translation that was available in the literature [15] for evaluation. The left side of Fig. 1 schematically illustrates the cGAN we implemented. We used the UT-Zap50K data set of shoe/handbag images and their corresponding line sketch data set [31]–[33] to train and test the implemented cGAN. As illustrated in the right of Fig. 1, it showed excellent performance as reported in the literature. However, we found that it was not robust against the incomplete or corrupted images given as the test input. This is illustrated in Figs. 2 and 3, where, unlike its high-quality performance with complete and corruption-free input images, the cGAN approach failed to produce the translation that we were looking for when it was tested with incomplete and corrupted input images. Specifically, the cGAN approach was not able to compensate for a

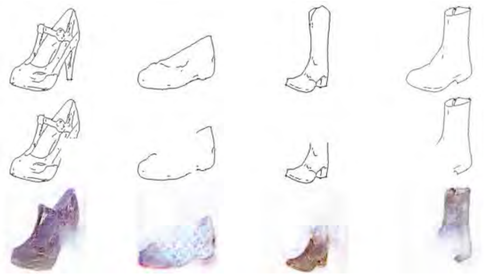


FIGURE 2. cGAN based image-to-image translation applied to incomplete images (Top: Complete sketches given as references, Middle: Incomplete sketches given as input images, Bottom: Translated output images). Notice that the translated output images inherit the missing parts of the given input sketches.

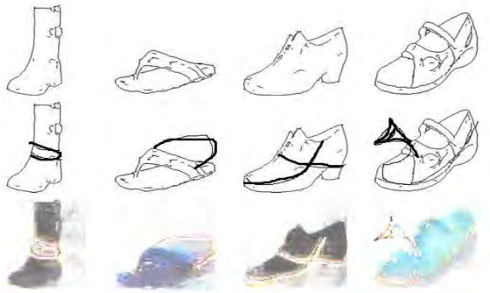
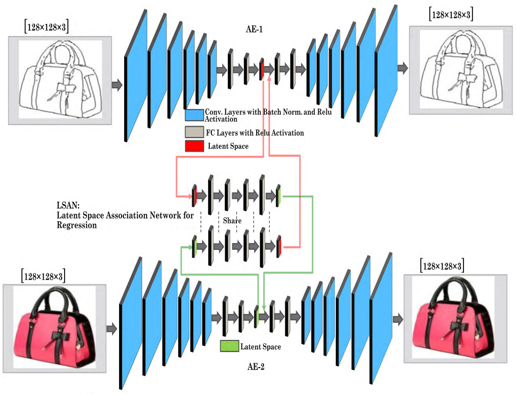


FIGURE 3. cGAN based image-to-image translation applied to corrupted images. (Top: Complete sketches given as references, Middle: Corrupted sketches given as input images, Bottom: Translated output images). Notice that the translated output images inherit the corrupted parts of the given input sketches.

missing or noisy part of the given input images to produce the typical, complete, and noise-free output images that we expected. This could have been because cGAN emphasizes learning the exact joint probability distribution of the image pairs given as the training data set rather than their generalization. As seen in Figs. 2 and 3, cGAN tends to incorporate the variation of an input into that of the corresponding output to emphasize accuracy in pair-wise association, but only at the expense of generality. Therefore, robust image-to-image translation that achieves a high degree of generality while keeping accuracy high remains a key issue to be solved. In addition, it would be desirable if we could establish a single unified framework with the capability of bidirectional translation that can directly handle applications that share a common ground of image translation, including image coloring, styling, de-noising, modification, and completion.

III. PROPOSED APPROACH

As a solution for the problem described above, we propose dual auto-encoder with bidirectional latent space regression or Bi-directionally Associative DualAE for short. The proposed BA-DualAE is configured with two auto-encoders the individual latent spaces of which are tightly associated by a bidirectional regression network. Once the auto-encoders are trained independently for their respective domains, the bidirectional regression network is trained to learn the general associations between data pairs. The proposed approach aims at achieving a sufficient level of generality while keeping accuracy high, for dealing with untrained, incomplete and corrupted data. This is based on the capability of an auto-encoder to abstract the input data into its latent space and of a regression network to learn the general association between the two latent space representations. With the capability of robust and bidirectional image translation, the proposed BA-DualAE is able to perform direct image completion with no iterative search, paving a way to a unified framework in image translation.


[128×128×3]
AE-1
[128×128×3]
Conv. Layers with Batch Norm. and Relu Activation
FC Layers with Relu Activation
Latent Space
LSAN:
Latent Space Association Network for Regression
Share
[128×128×3]
[128×128×3]
Latent Space
AE-2

Encoder Conv. Layers: [64×64×64], [16×16×196], [8×8×512], [4×4×512], [2×2×1024], [1×1×2000]
Encoder FC Layers: [800,500]
Latent Space: [200]
The structure of Decoder is exactly the reverse order of that of Encoder.
LSAN: [200,500,300,300,500,200]

FIGURE 4. The proposed robust cross-domain image to image translation network. The networks, AE-1 and AE-2, are trained to generate their output images same as their corresponding input images in their own domains, while LSAN is trained to regress the cross-domain relationship between the two latent spaces of AE-1 and AE-2. Note that the proposed architecture with LSAN allows a bidirectional cross-domain translation.

IV. NETWORK ARCHITECTURE

The proposed BA-DualAE consists of two deep auto-encoders (AEs) and a latent space association network (LSAN), as shown in Fig. 4. AEs are to map their input distributions into their respective latent space domains, while the LSAN is to define the cross-domain relationships by regressing the mapping between the two latent spaces from their respective AEs.

A. DEEP AUTO-ENCODERS

The first deep auto-encoders, AE-1, is trained based on the image data set of its own domain, $\mathcal{Y}_1$, such that it transforms an RGB image, $y_1 \varepsilon \mathcal{Y}$, given as an input into the corresponding latent space point,$q_1 (z_1 \mid y_1)$, of a lower dimension by its encoder, $Enc_1 (y_1)$, while reconstructing the latent space representation back to a close proxy, $y_1^{'}$, to the given input image by its decoder, $Dec_1(z_1|y_1)$. Formally, the latent space of AE-1 is described as

$$q_1 (z_1 \mid y_1) = Enc_1 (y_1) , \tag{1}$$

$q_1(z_1|y_1)$ is then transformed back to the reconstructed input image $y_1^{'}$ by the decoder of AE-1.

$$y_1^{'} = Dec_1 (z_1 \mid y_1) , \tag{2}$$

Similarly, AE-2 takes an RGB image, $y_2\ \mathcal{Y}$, as an input and maps it to the corresponding latent space point, $q_2(z_2|y_2)$, by its encoder, $Enc_2(y_2)$. The decoder takes $q_2\,(z_2\,|\,y_2)$ and reconstructs the input image $y_2^{'}$ as follows:

$$q_2\,(z_2\,|\,y_2) = Enc_2\,(y_2) \tag{3}$$

$$y_2^{'} = Dec_2\,(z_2\,|\,y_2)\,, \tag{4}$$

The following L_2 regression losses are used to train AE-1 and AE-2 for their respective domains:

$$L_{AE1} = \|y_1 - Dec_1(z_1|y_1)\|\,,\ y_1\mathcal{Y}_1 \tag{5}$$

$$L_{AE2} = \|y_2 - Dec_2(z_2|y_2)\|\,,\ y_2\mathcal{Y}_2 \tag{6}$$

The same architecture is used for AE-1 and AE-2 with the same numbers of layers and parameters; both have 16 layers with the input RGB images having resolution of 128×128×3. The number of filters is doubled up to the fourth layer and in the remaining four layers, the filters are doubled after two consecutive layers. We used stride of 2 in each layer and there is no pooling layer in our framework. In each layer, the convolution is followed by batch normalization [34] layer and activation layer. We used the ReLU [35] activation function in each of these layers except the last layer, where we used tanh activation function.

B. LATENT SPACE ASSOCIATION NETWORK FOR REGRESSION

For bidirectional image translation, we used LSAN, a fully connected association network that makes the correspondence between the projected latent space distributions of both the domains. In order to a build cross-domain relationship between the two AEs, LSAN makes the association between the marginal distribution of domain 2, projected by AE-2 with domain 1 marginal distribution projected by AE1. LSAN takes projected latent space from AE-1 and transforms it to the latent space of AE-2, where it then acts as input to the decoder of AE-2 for cross-domain translation from domain 1 to domain 2. Similarly, to transform from domain 2 to domain 1, LSAN takes the projected latent space from AE-2 and transforms it into the latent space of AE-1. The combined loss of LSAN is:

$$L_{LSAN} = q_1\,(z_1\,|\,y_1) - LSAN\,(q_2\,(z_2\,|\,y_2)) + q_2\,(z_2\,|\,y_2) - LSAN\,(q_1\,(z_1\,|\,y_1))\,. \tag{7}$$

where $q_1\,(z_1\,|\,y_1)$ and $q_2\,(z_2\,|\,y_2)$ are the encoder output of AE-1 and AE-2 respectively. LSAN is the association between two latent representations of both the domains. Both the loss function of the same domain and the cross-domain translation are minimized using the stochastic gradient method.

The combined loss of AE-LSAN is:

$$L_{\mathrm{AE}-LSAN} = L_{LSAN} + L_{AEs}. \tag{8}$$

V. TRAINING DETAILS

The training of BA-DualAE is based on two steps: The first step of training focuses on same-domain translation, where AEs are trained separately in their own domains. Once the translation in the same domain is complete, the next step is to associate the latent spaces of individual domains. In this step, we do cross-domain image translation by training LSAN; LSAN associates the latent spaces of two domains bidirectionally by minimizing the regression loss between its output the ground truth latent distribution. Here, the input to LSAN is the latent space representations from the encoders of AE-1 and AE-2, and its output is the latent space representations cross-transformed from one domain to another. After LSAN is trained, the latent space outputs from LSAN are connected to the respective decoders of AE-1 and AE-2. During the training, we used different data sets with different numbers of training samples, and we also used different mini-batch sizes for different data sets: For the small data sets, we used a smaller mini-batch and a larger mini-batch for larger data sets. Considering that in a highly nonlinear space, a high or a low learning rate may either overshoot the desired minima or keep the network lingering around the local minima, we adopted the learning rate of 0.000025 for updating the parameters of both AEs and LSAN; we also used Adam [36] to optimize the network parameters with beta = 0.5.

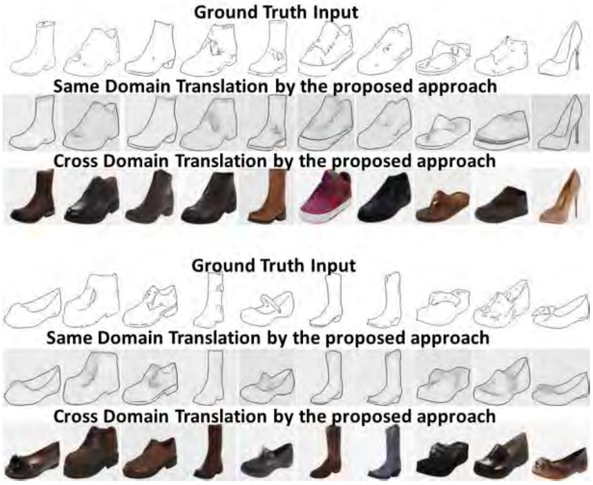

Ground Truth Input
Same Domain Translation by the proposed approach
Cross Domain Translation by the proposed approach
Ground Truth Input
Same Domain Translation by the proposed approach
Cross Domain Translation by the proposed approach

FIGURE 5. Analysis of the proposed approach for same as well as cross-domain image to image translation. The sketch-based input testing samples are translated to sketched based samples in the same domain as well as to its corresponding real representation in the cross-domain.

VI. EXPERIMENTS

First, we assessed the performance of BA-DualAE in terms of the bidirectional translation of images from one domain to its corresponding cross-domain and vice versa; this was based on pairwise data of natural images and their sketches from two data sets: UT-Zap50K [31] and the handbag set [33]. UT-Zap50K contains about 50k samples of different types of shoes with 200 separate testing samples; the handbag data set contains about 138k samples with 200 samples separated as testing samples. The edge-based representations for both data sets are obtained based on the Holistically Edge Detection (HED) algorithm [32]. The experimental results are shown in Figs. 5 to 12, and the detailed analysis of

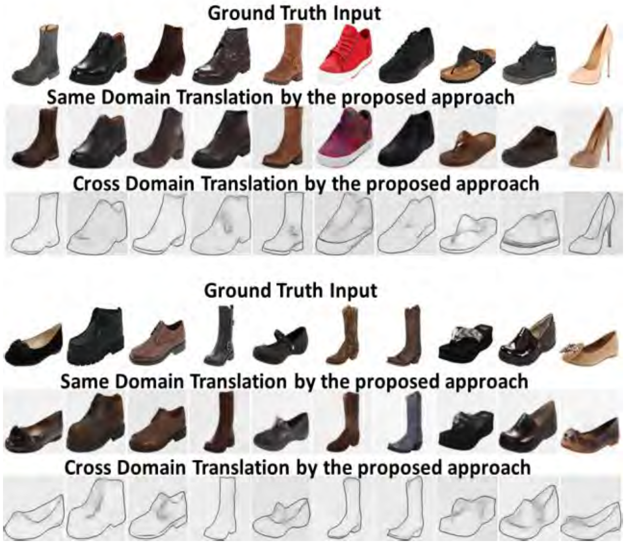

Ground Truth Input
Same Domain Translation by the proposed approach
Cross Domain Translation by the proposed approach
Ground Truth Input
Same Domain Translation by the proposed approach
Cross Domain Translation by the proposed approach

FIGURE 6. Analysis of the proposed approach for same as well as cross-domain image to image translation. The real representation-based input testing samples are translated to real samples in the same domain as well as to its corresponding sketch-based representation in the cross-domain.

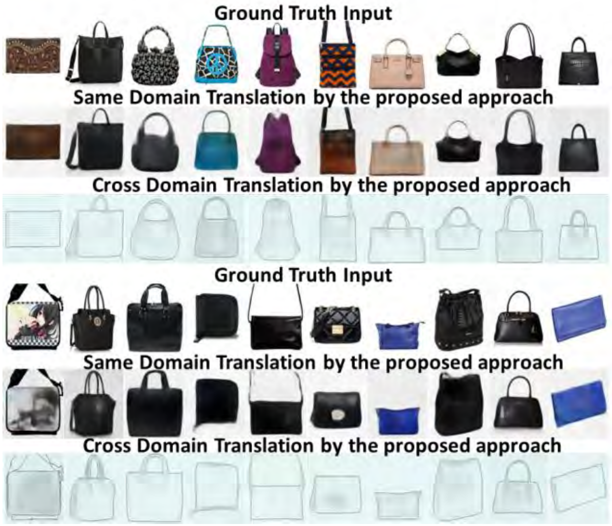

Ground Truth Input
Same Domain Translation by the proposed approach
Cross Domain Translation by the proposed approach
Ground Truth Input
Same Domain Translation by the proposed approach
Cross Domain Translation by the proposed approach

FIGURE 7. Analysis of the proposed approach for the same as well as cross-domain image to image translation. The real representation based input testing samples are translated to real samples in the same domain as well as to its corresponding sketch-based representation in the cross-domain.

these results is given in Section 5.1. Second, we evaluated BA-DualAE in terms of its capability of direct image completion as a means of providing an insight into data interpretations and generality in learning. The experimental results are shown in Figs. 13 to 16, and the detailed analysis of these results is given in Section 5.2. It is noted that we analyzed the experimental results not only qualitatively but also quantitatively by defining the mean square error from the ground truth to the translated images, as shown in TABLE 1 and 2. During the experimental setup, we used RGB images with 128 × 128 × 3 as the resolution.

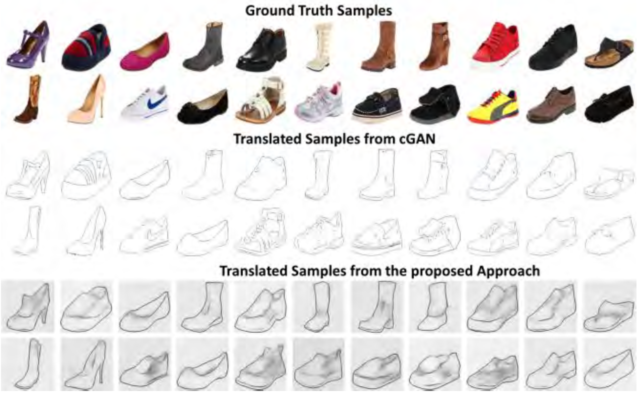

Ground Truth Samples
Translated Samples from cGAN
Translated Samples from the proposed Approach

FIGURE 8. Comparative analysis of the proposed approach with cGAN [15] based on the complete and uncorrupted data as shown.

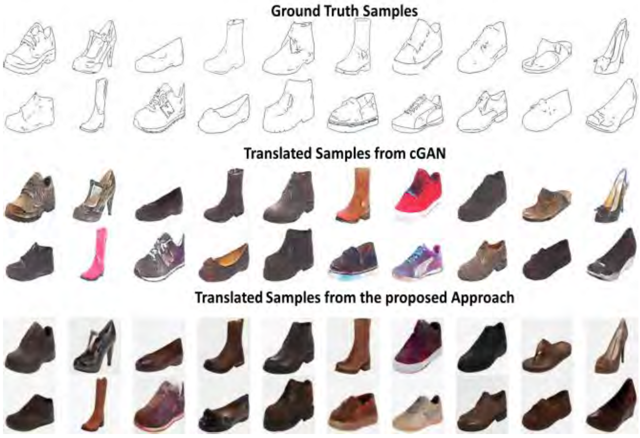

Ground Truth Samples
Translated Samples from cGAN
Translated Samples from the proposed Approach

FIGURE 9. Comparative analysis of the proposed approach with cGAN [15] based on the complete and uncorrupted data as shown.

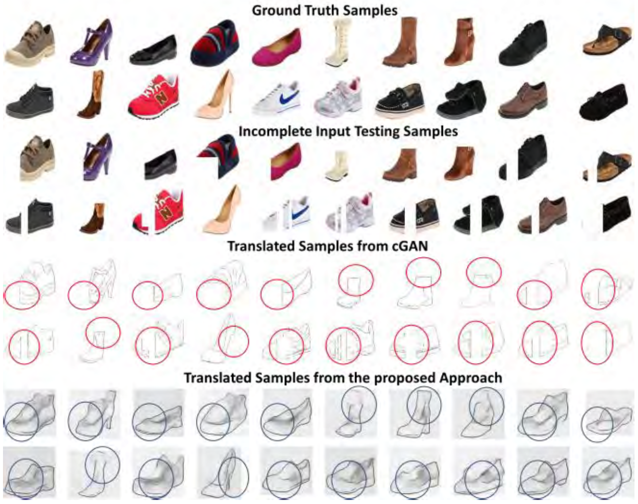

Ground Truth Samples
Incomplete Input Testing Samples
Translated Samples from cGAN
Translated Samples from the proposed Approach

FIGURE 10. Comparative analysis of the proposed approach with cGAN [15] based on the incomplete data as shown.

A. CROSS-DOMAIN IMAGE-TO-IMAGE TRANSLATION

Shoes-to-Sketch Translation: In this experiment, we used the UT-Zap50K data set along with their corresponding sketch-based representations for training. Figure 5 shows the results of translation from the sketch to shoes with

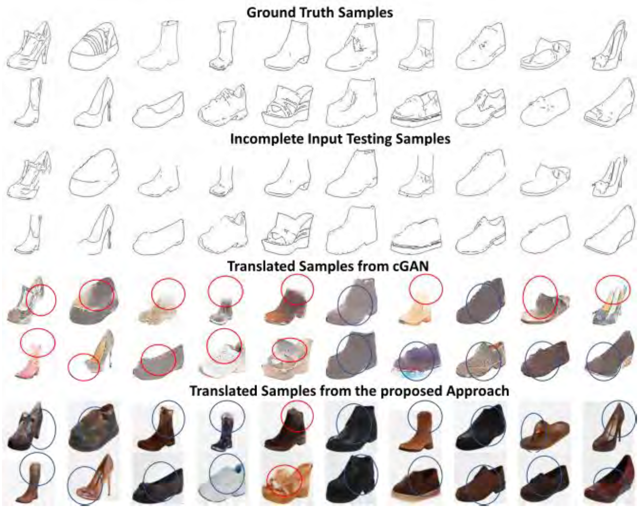

Ground Truth Samples
Incomplete Input Testing Samples
Translated Samples from cGAN
Translated Samples from the proposed Approach

FIGURE 11. Comparative analysis of the proposed approach with cGAN [15] based on the incomplete data as shown.

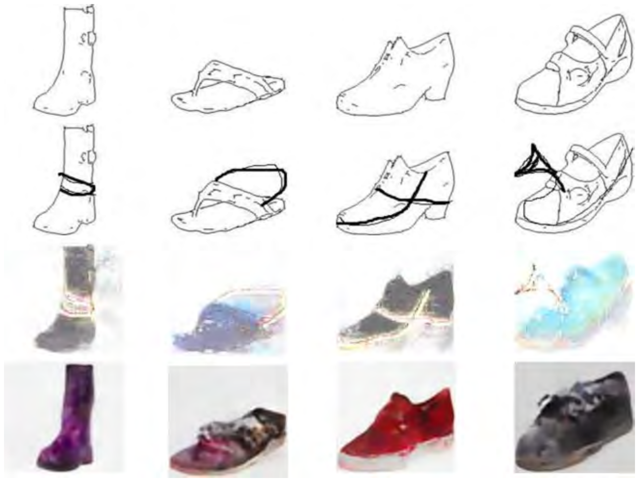


FIGURE 12. Comparative analysis of the proposed approach with cGAN [15] based on the corrupted data as shown.

the sketch-based input testing samples as inputs. The input testing samples were translated to within-domain sketches as well as to their cross-domain real images, as shown in Figure 5. Fig. 6 shows the within- and cross-domain translations from the real images to their corresponding real images and sketches. Figs. 5 and 6 illustrate that the proposed BA-DualAE produced realistic within-domain and cross-domain image translations with high accuracy.

Handbags-to-Sketch Translation: We repeated the experiments for UT-Zap50K for the handbag dataset. The within- and cross-domain results are illustrated in Fig. 7. Similar to the case with UT-Zap50K, the proposed BA-DualAE produced realistic within- and cross-domain image translations with high accuracy for the handbag dataset.

1) COMPARATIVE ANALYSIS WITH COMPLETE AND UNCORRUPTED DATA

We performed a comparative analysis of BA-DualAE with cGAN [15] for cross-domain image translation. The results are shown in Figs. 8 and 9, where Fig. 8 presents the results of translation from sketch to real image while Fig. 9 shows real image to sketch for both cGAN and BA-DualAE. For the comparative analysis, we trained both cGAN and BA-DualAE on the UT-Zap50K data set and its corresponding edge representation. We set the batch size to 32 in both cases.

The ground truth of real images and sketches, illustrated in Figs. 8 and 9, was used to assess accuracy. Note that the capability of BA-DualAE for bidirectional translation allows it to carry out both direction and translation without additional training. The results indicate that both cGAN

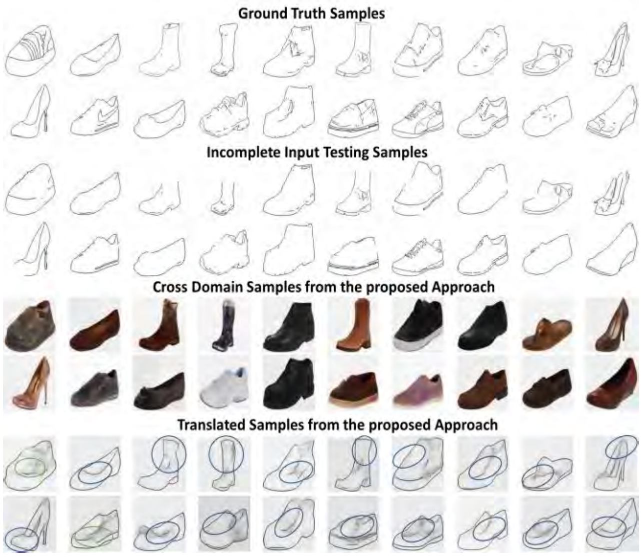

Ground Truth Samples
Incomplete Input Testing Samples
Cross Domain Samples from the proposed Approach
Translated Samples from the proposed Approach

FIGURE 13. Illustration of our models on image completion by utilizing the cross-domain correspondence between the images. The ground truth samples are cropped at random locations and the incomplete input testing samples are then translated to the corresponding cross-domain and same domain respectively.

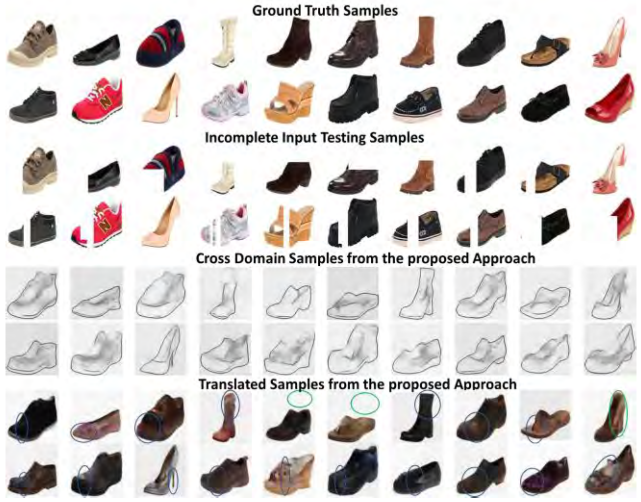

Ground Truth Samples
Incomplete Input Testing Samples
Cross Domain Samples from the proposed Approach
Translated Samples from the proposed Approach

FIGURE 14. Illustration of our models on image completion by utilizing the cross-domain correspondence between the images. The ground truth samples are cropped at random locations and the incomplete input testing samples are then translated to the corresponding cross domain and same domain respectively.

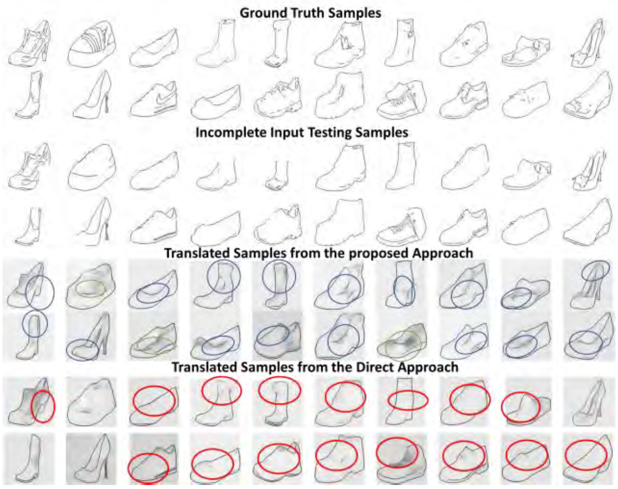

Ground Truth Samples
Incomplete Input Testing Samples
Translated Samples from the proposed Approach
Translated Samples from the Direct Approach

FIGURE 15. Comparative analysis of the proposed approach with direct image completion approach. The ground truth samples are cropped at random locations and the incomplete input testing samples are then translated for image completion using the proposed approach and the direct image completion approach.

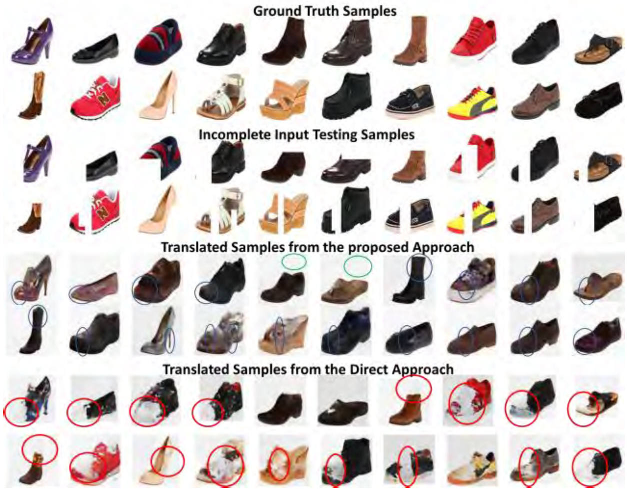

Ground Truth Samples
Incomplete Input Testing Samples
Translated Samples from the proposed Approach
Translated Samples from the Direct Approach

FIGURE 16. Comparative analysis of the proposed approach with direct image completion approach for image completion. The ground truth samples are cropped at random locations and the incomplete input testing samples are then translated for image completion using the proposed approach and direct image completion approach.

TABLE 1. The Mean square error based on UT-Zap50K and its edge-based representation.

UT-Zap50K +Edge Representation: No testing data distortion	
Approaches	MSE
cGAN	144.5
BA-DualAE	**117.6**

and BA-DualAE produce high accuracy in cross-domain translation with some interesting differences: For instance, cGAN preferred to maintain the fine details of the samples while BA-DualAE preferred maintaining global features. As mentioned previously, this clearly indicates the emphasis of cGAN on accuracy over generality whereas the proposed BA-DualAE focuses much more on generality while keeping accuracy high. Note that cGAN as we implemented it has skip connections for improved performance.

TABLE 2. Mean square error based on reconstructing partially occluded samples using UT-Zap50K and its edge-based representation.

Approaches	Low Level of occlusion	High Level of occlusion
cGAN	101.222	127.521
BA-DualAE	**76.202**	**109.302**

TABLE 3. Mean square error and SSIM based on reconstructing incomplete samples using UT-Zap50K and its edge-based representation.

Approaches	MSE	SSIM
cGAN	30.378	0.605
AE	26.769	0.621
BA-DualAE	**22.54**	**0.644**

TABLE 1 shows the quantitative performance analysis of cGAN and BA-DualAE based on the mean square error from the ground truth. The error is calculated based on the first 100 samples of the UT-Zap50K test data set and their corresponding sketches. TABLE 1 shows that the proposed BA-DualAE has about 20% advantage over cGAN based on the mean square error. To analyze the robustness, we conducted the same performance test with the input test samples intentionally modified by occlusion; TABLE 2 shows the result, that the proposed BA-DualAE has again a 20% advantage over cGAN. However, in the cases of the occluded parts, indicated by circles in Figs.10 and 11, BA-DualAE showed much greater capacity to recover the missing parts and greater robustness. During this analysis, we also observed that the proposed approach successfully recovers the shape to a certain point of occlusion but that when the occlusion is heavy, the occluded samples are transformed to the nearby samples in the database, demonstrating the generality of BA-DualAE.

We also evaluated the proposed network in terms of the mean square error and structural similarity index (SSIM) for randomly selected two batches of incomplete testing samples as shown in TABLE 3. The proposed approach outperformed cGAN [15] and AE [21]. The SSIM shows the structural similarity between the data generated by the network and their corresponding ground truth samples; the SSIM was higher with the proposed approach in terms of the incomplete input testing data than with AE [21] and cGAN [15].

2) COMPARATIVE ANALYSIS WITH INCOMPLETE AND CORRUPTED DATA

We conducted the qualitative evaluation of BA-DualAE and cGAN based on incomplete and corrupted data (Figs. 10 and 11). In Fig. 10, parts of the real images (the first and the second rows) from UT-Zap50K dataset are removed at random locations and used as the input samples for testing (the third and the fourth rows). On the other hand, in Fig. 11, parts of sketches are removed out at random locations and then used as the input samples for testing. The results clearly show the advantage of the proposed approach over cGAN based approach. The cGAN based approach carries the missing parts from the input images to the translated output images, as illustrated by the red circles. In contrast, the BA-DualAE successfully translates the complete images by recovering the missing parts in the input images, as illustrated by the blue circles.

Furthermore, as shown in Fig. 12, we also corrupted the input testing data by adding noise at random locations, as illustrated by darker lines. Then, we input the corrupted data to cGAN based approach and to BA-DualAE for comparative analysis. Fig. 12 shows clearly that BA-DualAE is far more robust to the corruption in the input testing samples than was cGAN: BA-DualAE successfully removed the unnecessary corruption in the data, whereas cGAN based approach reflected the corrupted portions in the output images. The comparative analysis with incomplete and corrupted data indicates clearly that the proposed BA-DualAE approach is more robust than the cGAN based approach. As briefly mentioned before, the cGAN based approach puts its emphasis more on accuracy due to the nature of cGAN, whereas the proposed BA-DualAE approach exploits the respective capabilities of LSAN and DualAE for the generalization in association and the accuracy in the construction.

B. IMAGE COMPLETION

Image completion aims at completing the missing pixels in an image with a most likely pixel configuration for the surrounding context; we need an efficient yet robust way of estimating a correct pixel configuration, preferably without explicit training with incomplete data. We found that we could take advantage of the robust and bidirectional image translation capability of the proposed BA-DualAE to make direct yet robust image completion possible in a two-step process: 1) input an incomplete, partially occluded or corrupted image to be completed to an AE of BA-DualAE in the same domain as the input and obtain the corresponding cross-domain output image and 2) input the cross-domain output image obtained from 1) to another AE of BA-DualAE in the same domain as the output image from 1) and obtain the corresponding cross-domain output image. Then, the resulting output image from 2) represents the completed image we target. The above two-step image completion process has a clear advantage in image completion in that during the first pass, the network estimates the most likely configuration of the missing pixels while during the second pass, the network makes use of the most likely configuration of missing pixels from the first pass to further refine the configuration of missing pixels.

To evaluate the proposed direct two-step image completion based on BA-DualAE, we conducted the following experiments: First, we modified the input testing samples by removing parts of the images at random locations; we then used the modified samples are then used as input to BA-DualAE to follow the two-step process described above. Figs. 13 and 14 illustrate the intermediate and final outputs from the two-step image completion process that we applied for the sketch domain (Fig. 13) and the real domain (Fig. 14) image completion. The figures show clearly how the incomplete samples originally input into the network are transformed to the corresponding cross-domain samples and, eventually, transformed back to the original domain as the complete samples. Note that the recovered missing pixels in the incomplete samples are highlighted by blue and green circles; the blue circles represent the actual recovery of the missing pixels, and the green circles represent the additional pixels the network generated as the missing pixels. The reason for this recovery of additional pixels is the generality of the proposed approach in image translation: When the corrupted samples resemble several other samples in the training data, the network tends to generalize the image translation by accommodating those samples. For the same reason, we also observed that the heavily corrupted samples were transformed into different samples that more resemble the corrupted samples. Note that because our proposed framework is bidirectional, it is straightforward to change the domain of incomplete input images in order to suit applications.

Furthermore, to evaluate the proposed image completion objectively, we performed a comparative analysis of the proposed image completion against the direct shape completion [21]; by direct image completion, we mean using AE-1 or AE-2 separately for the image completion in its domain without going through LSAN. The experimental results based on the sketch and real image data are shown in Figs. 15 and 16, respectively; the figures show clearly that the proposed framework outperforms the direct image completion based on AE. The direct image completion failed to complete the missing pixels in most of the examples as marked by the red circles in Figs. 15 and 16.

Note that there are some examples for which the direct image completion produces better results in color representation, although it lacks considerably in generality compared with the proposed approach.

VII. CONCLUSION AND DISCUSSION

We presented a bidirectionally associative dual autoencoder with a latent space regression network—in short, BA-DualAE—to solve the problem of robustness in conventional approaches. As we demonstrated, the proposed BA-DualAE is highly robust against untrained, incomplete, or corrupted data, resulting in a high degree of generality while keeping accuracy high. As far as the mean square error representing a rough measure of robustness, BA-DualAE is 20% better than conventional approaches, particularly showing marked performance advantages around incomplete or corrupted parts. Furthermore, the capability of BA-DualAE for robust and bidirectional translation enables it to perform direct image completion without an iterative search or explicit training. It is noted that the proposed network can be applied to various within-domain and cross-domain image translation tasks, is extensible to unsupervised learning, and is easily configurable into distributed AEs. These extensions will be best left as future works.

ACKNOWLEDGMENT

(Sukhan Lee and Naeem Ul Islam are co-first authors.)

REFERENCES

[1] Y. Taigman, A. Polyak, and L. Wolf. (2016). ''Unsupervised cross-domain image generation.'' [Online]. Available: https://arxiv.org/abs/1611.02200

[2] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, ''Learning to discover cross-domain relations with generative adversarial networks,'' in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1857–1865.

[3] M.-Y. Liu, T. Breuel, and J. Kautz, ''Unsupervised image-to-image translation networks,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 700–708.

[4] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, ''Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8789–8797.

[5] J.-Y. Zhu *et al.*, ''Toward multimodal image-to-image translation,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 465–476.

[6] N. Ul Islam and S. Lee, ''Cross domain image transformation using effective latent space association,'' in *Proc. Int. Conf. Intell. Auton. Syst.* Cham, Switzerland: Springer, 2018, pp. 706–716.

[7] C. Ledig *et al.*, ''Photo-realistic single image super-resolution using a generative adversarial network,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.

[8] L. A. Gatys, A. S. Ecker, and M. Bethge, ''Image style transfer using convolutional neural networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2414–2423.

[9] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras, ''Neural face editing with intrinsic image disentangling,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5541–5550.

[10] Y. Zhang, B. Du, and L. Zhang, ''A sparse representation-based binary hypothesis model for target detection in hyperspectral images,'' *IEEE Trans. Geosci. Remote Sens.* vol. 53, no. 3, pp. 1346–1354, Mar. 2015.

[11] B. Du, Y. Zhang, L. Zhang, and D. Tao, ''Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images,'' *IEEE Trans. Image Process.* vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[12] B. Du and L. Zhang, ''A discriminative metric learning based anomaly detection method,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.

[13] B. Du and L. Zhang, ''Target detection based on a dynamic subspace,'' *Pattern Recognit.*, vol. 47, no. 1, pp. 344–358, 2014.

[14] I. Goodfellow *et al.*, ''Generative adversarial nets,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, ''Image-to-image translation with conditional adversarial networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.

[16] N. Ul Islam and S. Lee, ''Interpretation of deep CNN based on learning feature reconstruction with feedback weights,'' *IEEE Access*, vol. 7, pp. 25195–25208, 2019.

[17] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, ''Learning from simulated and unsupervised images through adversarial training,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2107–2116.

[18] Z. Yi, H. Zhang, P. Tan, and M. Gong, ''DualGAN: Unsupervised dual learning for image-to-image translation,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2849–2857.

[19] Z. Cheng, Q. Yang, and B. Sheng, ''Deep colorization,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 415–423.

[20] S. Iizuka, E. Simo-Serra, and H. Ishikawa, ''Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification,'' *ACM Trans. Graph.*, vol. 35, no. 4, 2016, Art. no. 110.

[21] P. Baldi, ''Autoencoders, unsupervised learning, and deep architectures,'' in *Proc. ICML Workshop Unsupervised Transf. Learn.*, 2012, pp. 37–49.

[22] D. P. Kingma and M. Welling. (2013). ''Auto-encoding variational bayes.'' [Online]. Available: https://arxiv.org/abs/1312.6114

[23] D. J. Rezende, S. Mohamed, and D. Wierstra. (2014). ''Stochastic backpropagation and approximate inference in deep generative models.'' [Online]. Available: https://arxiv.org/abs/1401.4082

[24] Y. Li, K. Swersky, and R. Zemel, ''Generative moment matching networks,'' in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1718–1727.

[25] A. Van den Oord, A. N. Kalchbrenner, L. Espeholt, O. Vinyals, and A. Graves, ''Conditional image generation with PixelCNN decoders,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4790–4798.

[26] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski, ''Plug & play generative networks: Conditional iterative generation of images in latent space,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4467–4477.

[27] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus, ''Deep generative image models using a Laplacian pyramid of adversarial networks,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1486–1494.

[28] A. Radford, L. Metz, and S. Chintala. (2015). ''Unsupervised representation learning with deep convolutional generative adversarial networks.'' [Online]. Available: https://arxiv.org/abs/1511.06434

[29] M. Arjovsky, S. Chintala, and L. Bottou. (2017). ''Wasserstein GAN.'' [Online]. Available: https://arxiv.org/abs/1701.07875

[30] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, ''Context encoders: Feature learning by inpainting,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2536–2544.

[31] A. Yu and K. Grauman, ''Fine-grained visual comparisons with local learning,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 192–199.

[32] S. Xie and Z. Tu, ''Holistically-nested edge detection,'' in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1395–1403.

[33] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, ''Generative visual manipulation on the natural image manifold,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 597–613.

[34] S. Ioffe and C. Szegedy. (2015). ''Batch normalization: Accelerating deep network training by reducing internal covariate shift.'' [Online]. Available: https://arxiv.org/abs/1502.03167

[35] V. Nair and G. E. Hinton, ''Rectified linear units improve restricted Boltzmann machines,'' in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010.

[36] D. P. Kingma and J. Ba. (2014). ''Adam: A method for stochastic optimization.'' [Online]. Available: https://arxiv.org/abs/1412.6980

SUKHAN LEE received the B.S. and M.S. degrees in electrical engineering from Seoul National University, South Korea, in 1972 and 1974, respectively, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, USA, in 1982.

From 1983 to 1997, he was with the Department of Electrical Engineering, University of Southern California, also with the Department of Computer Science, University of Southern California, and also with the Jet Propulsion Laboratory, California Institute of Technology, as a Senior Member of Technical Staff, from 1990 to 1997. From 1998 to 2003, he was an Executive Vice President and a Chief Research Officer with the Samsung Advanced Institute of Technology. He has been a Professor of information and communication engineering and has been a WCU Professor of interaction science with Sungkyunkwan University, since 2003. He was designated the Dean of the Graduate School, Sungkyunkwan University, in 2011. He is also serving as the Director of the Intelligent Systems Research Institute. His research interests include cognitive robotics, intelligent systems, and micro/nano electro-mechanical systems.

Dr. Lee is currently a fellow of the Korean National Academy of Science and Technology.

NAEEM UL ISLAM received the B.S. degree in electrical and electronics engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2008. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Sungkyunkwan University, Suwon, South Korea.

His research interests include artificial intelligence (AI), machine learning, computer vision, and deep learning.