Generative Adversarial Capsule Network With ConvLSTM for Hyperspectral Image Classification

Wei-Ye Wang, Heng-Chao Li, *Senior Member, IEEE*, Yang-Jun Deng, Li-Yang Shao, *Senior Member, IEEE*, Xiao-Qiang Lu, *Senior Member, IEEE*, and Qian Du, *Fellow, IEEE*

***Abstract*— Recently, deep learning has been widely applied in hyperspectral image (HSI) classification since it can extract high-level spatial–spectral features. However, deep learning methods are restricted due to the lack of sufficient annotated samples. To address this problem, this letter proposes a novel generative adversarial network (GAN) for HSI classification that can generate artificial samples for data augmentation to improve the HSI classification performance with few training samples. In the proposed network, a new discriminator is designed by exploiting capsule network (CapsNet) and convolutional long short-term memory (ConvLSTM), which extracts the low-level features and combines them together with local space sequence information to form the high-level contextual features. In addition, a structured sparse $L_{2,1}$ constraint is imposed on sample generation to control the modes of data being generated and achieve more stable training. The experimental results on two real HSI data sets show that the proposed method can obtain better classification performance than the several state-of-the-art deep classification methods.**

***Index Terms*— Capsule network (CapsNet), convolutional neural network (CNN), data augmentation, deep learning, generative adversarial network (GAN), hyperspectral image (HSI) classification.**

I. Introduction

HYPERSPECTRAL images (HSIs) contain abundant spectral information in hundreds of spectral bands for each image pixel. With improved spatial resolution, HSIs can also present rich contextual structure information of imaged scenes. Generally, different land-cover materials have different spectral features, which can be used to distinguish ground objects. Therefore, HSIs have been widely applied in many areas [1], [2], such as agricultural monitoring, military surveillance, and environmental protection.

Over the last few decades, HSI classification is an increasingly hot research topic. Numerous classification methods, such as random forests [3], logistic regression [4], and support vector machine (SVM) [5]–[7], have been developed, which usually employ handcrafted features in a low-dimensional space. In recent years, since deep learning methods can extract high-level features automatically, they have been successfully applied for HSI classification and achieve outstanding classification performance [8]. For example, Chen *et al.* [8] proposed a 3-D convolutional neural network (CNN)-based deep framework, which adopted 3-D kernels to extract the spectral–spatial features for HSI classification. Li *et al.* [9] utilized a fully connected CNN to extract deep features of HSIs and subsequently employed an extreme learning machine (ELM) for classification rather than the end-to-end training. A cascaded recurrent neural network (RNN) model was proposed by Hang *et al.* [10], which utilizes gated recurrent units to explore the redundant and complementary information of HSIs. These works illustrated the successful applications of deep learning in the HSI classification.

However, the deep learning methods require vast amounts of data to train a stable model for feature extraction of HSIs; otherwise, their performance will be limited. As well known, data annotation is very difficult and expensive for HSIs. The generative adversarial network (GAN) [11] proposed recently could provide a good solution to the lack of labeled data in an HSI analysis, which captures the data distribution and generates the simulated samples to augment the data sets. For instance, He *et al.* [12] proposed a semisupervised HSI classification method based on GAN, which is achieved by adding samples from the generator to the features and increasing the dimension of the classifier's output. A new GAN-based HSI classification framework was proposed by Zhu *et al.* [13], which utilizes the theory of an auxiliary classifier GAN (AC-GAN) [14] and the CNN model to improve the final classification performance.

Although the GAN has shown its promising effectiveness from a variety of aspects, it also suffers from instability issues due to its adversarial idea [11]. To address the small-size sample issue, we propose a novel GAN, which can generate artificial samples for data augmentation to improve the classification performance. Specifically, a new loss function is defined by introducing a tradeoff parameter to balance the

Manuscript received November 14, 2019; revised January 21, 2020; accepted February 20, 2020. Date of publication March 11, 2020; date of current version February 25, 2021. This work was supported by the National Natural Science Foundation of China under Grant 61871335. *(Corresponding authors: Heng-Chao Li; Yang-Jun Deng.)*

Wei-Ye Wang, Heng-Chao Li, and Yang-Jun Deng are with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: lihengchao_78@163.com; dyj2012@yeah.net).

Li-Yang Shao is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China.

Xiao-Qiang Lu is with the Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, Xi'an 710119, China, and also with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China.

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA.

Color versions of one or more of the figures in this letter are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2020.2976482

label and source losses, which can generate global coherent samples rather than specific coherent samples. Meanwhile, a structured sparse $L_{2,1}$ constraint is further added, which imposes a control on the mode and achieve a more stable training process. Moreover, a new discriminator is designed. On the one hand, capsule vector and dynamic routing mechanism [15] are employed to improve the feature representation ability and overcome the deficiency of position information loss in the existing CNN-based GAN models. On the other hand, considering that hyperspectral data is 3-D data cube that simultaneously has spatial structure and spectral sequence correlation, the convolutional long short-term memory (ConvLSTM) [16] is integrated into the discriminator for extracting this useful information. In addition, ConvLSTM can control the generator to impose the generated samples with contextual correlation and more similarity with real HSI data. By undertaking this integration, the proposed ConvLSTM capsule GAN (CCAPS-GAN) not only makes full use of generated artificial samples for data augmentation but also exploits more discriminative spatial–spectral features to improve the classification performance.

II. Review of GAN

GAN is a new model proposed by Goodfellow *et al.* [11], which contains a generative model G and a discriminative model D. The generator G aims to capture the potential distribution p_{data} of real data and learns a new distribution p_g of generative fake data, which is required to be as close as possible to the distribution p_{data}. The discriminator D is used to estimate the distance between p_g and p_{data}, which can determine whether the input samples are from the real data or not. G and D are simultaneously trained, and their parameters are alternately updated, as if they abide by the two-player min–max game with the value function $V(G, D)$. The objective function aims to solve the minimax problem represented as

$$\min_G \max_D V(G, D) = E_{x\sim p_{\text{data(x)}}}[\log D(x)] + E_{x\sim p_{\text{g(z)}}}[\log(1 - D(G(z)))] \quad (1)$$

where E is the expectation operator, x is the training data with the true data distribution p_{data}, and z represents the noise variable sampled from distribution p_g. Through the adversarial and competition fashion of two networks, the discriminator will proceed constantly and effectively in the training process, rather than getting trapped into overfitting frequently with limited training samples.

III. Proposed CCAPS-GAN Classification Framework

A. Proposed Loss Function

Despite the fact that AC-GAN [14] can be used in image classifications, its structure and loss function cause the generated samples having specific coherent with fake labels, which may degrade the classification performance. Here, we propose a novel classification framework by optimizing AC-GAN and the corresponding objective function is modified to consider the likelihood of the correct source or classes and to generate globally coherent samples rather than specific coherent samples. In the proposed network, every generated sample $x_{\text{fake}} = G(z)$ has a corresponding pseudolabel $c + 1$, which can be used for image classification. Specifically, the objective function can be written as

$$\begin{aligned}\min_G \max_D V(G, D) &= E_{x\sim p_{\text{data(x)}}}[\log D(x)] + E_{x\sim p_{\text{g(z)}}}[\log(1 - D(G(z)))] \\ &\quad + \lambda_1\big(E_{x\sim p_{\text{data(x)}}}[\log D(x, c)] \\ &\qquad + E_{x\sim p_{\text{g(z)}}}[\log(1 - D(G(z), c + 1))]\big) \quad (2)\end{aligned}$$

where λ_1 is a tradeoff parameter of label and source losses. D is optimized to maximize (2). The generated fake samples with pseudolabel can be treated as augmented data that increase the number of training samples and prevent overfitting. The generated samples and real samples are used together by the underlying classifier to improve predictions on data sets. Note that minimizing (2) with $\lambda_1 = 0$ can perform unsupervised learning by ignoring the component of the loss that arose from class labels. Hence, the fake data generated by G possess a more similar distribution with real data. The proposed loss function can be degenerated into the AC-GAN's loss function when D and G are optimized with $\lambda_1 = 1$ and $\lambda_1 = -1$. In order to obtain higher quality samples with global coherence, the more structure along with a special cost function is added to the GAN latent space. Then, the structured sparse $L_{2,1}$ constraint is adopted in the discriminator, which forces the discriminator to be more stable during the training process. Therefore, the final objective function can be formulated as

$$\min_G \max_D V(G, D) + \lambda_2 L_{2,1}(D) \quad (3)$$

where λ_2 is the parameter for $L_{2,1}$ regularization. By using the structured sparsity, the proposed model resolves the mode collapse problem of GAN and improves the convergence of target distribution. Therefore, the proposed model can impose a control on the mode of data being generated and achieve more stable training as well.

B. Proposed Discriminator

In order to obtain better classification performance, a new discriminator architecture is designed, which integrates the capsule network (CapsNet) [15] with ConvLSTM [16] to improve the discriminant property of the model. Specifically, the proposed discriminator is based on CapsNet that is composed of many capsule neurons, which overcomes the deficiencies of the existing CNN-based models. Each capsule contains the vectors of instantiation parameters instead of the classical scalar outputs, which characterize the input data with highly abstract features and their corresponding instantiation parameters. Meanwhile, the dynamic routing agreement is adopted between the PrimaryCaps and EntityCaps layers, which can reinforce correlation connections and reduce the uncorrelated connections. It not only retains all the underlying capsules but also considers the importance of each capsule, and therefore, the dynamic routing is more efficient than the maximum pooling. In this way, the features extracted by the discriminator become more informative, which can better characterize the HSI data.

However, only using original CapsNet as the discriminator will lead to the loss of spatial structure and sequence correlation, which are naturally existed in the hyperspectral data. Therefore, ConvLSTM is adopted in the discriminator to better preserve the intrinsic structure of hyperspectral data,

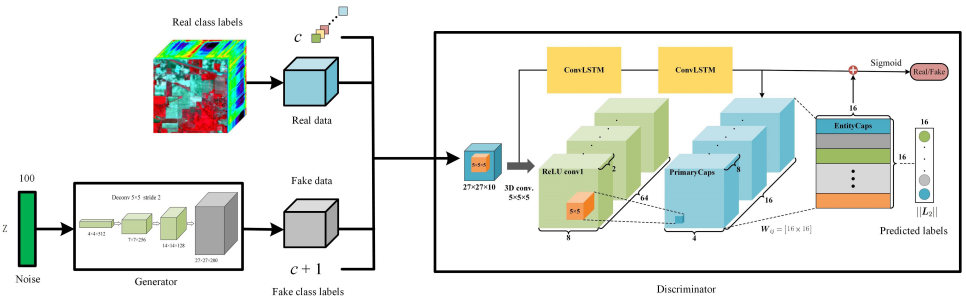

Fig. 1. Overall architecture of the proposed CCAPS-GAN framework.

TABLE I
OA UNDER DIFFERENT SIZES OF LOCAL WINDOW

$s \times s$	Indian Pines (OA)	$s \times s$	University of Houston (OA)
11×11	68.54	7×7	68.32
15×15	70.67	9×9	**70.69**
19×19	71.46	11×11	69.82
23×23	71.85	13×13	69.42
27×27	**72.31**	15×15	68.57
31×31	71.77	17×17	67.80

TABLE II
OA UNDER DIFFERENT FEATURE MAPS

M	Indian Pines (OA)	M	University of Houston (OA)
{8, 16}	71.21	{8, 16}	69.54
{16, 32}	**72.31**	{16, 32}	70.13
{32, 64}	71.39	{32, 64}	**70.69**
{64, 128}	70.98	{64, 128}	69.87

TABLE III
OA UNDER DIFFERENT SPARSE COEFFICIENTS

λ_2	Indian Pines (OA)	University of Houston (OA)
0	71.41	69.73
0.0005	71.43	70.18
0.001	**72.31**	**70.69**
0.005	71.12	69.33
0.01	69.56	68.17
0.05	68.21	66.87
0.1	65.44	65.32

and the special gate mechanisms enable it to extract the contextual features among the adjacent sequential data. As shown in Fig. 1, two-level ConvLSTM is deployed in the proposed discriminator. The final output of the high-level ConvLSTM layer is regarded as the final contextual features, which is used to construct primary capsule neurons. Thus, high-level capsule neurons (i.e., EntityCaps) can improve the classification performance by incorporating both sequence contextual information and spectral–spatial information. Moreover, the outputs of the high-level ConvLSTM and EntityCaps are together conveyed to the sigmoid classifier for capturing the distribution of real data. By undertaking this integration, we have some degree of control over the generator to impose the generated samples with contextual correlation and more similarity with real data.

Fig. 1 shows the architecture of the generator and discriminator in the proposed CCAPS-GAN framework. Specifically, CNN is still used as the generator, which accepts the noise as the input and generates artificial samples. The CCAPS-GAN is trained in the traditional way that the discriminator D and generator G are trained alternately. Then, the training of both G and D will converge when G can generate fake data that are most similar to real data and D cannot distinguish the fake data and real data.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the effectiveness of the proposed method, experiments are conducted on two public hyperspectral data sets, i.e., "Indian Pines" and "University of Houston". The Indian Pines data set with an image size of 145 × 145 was collected by the AVIRIS sensor over northwestern Indiana in USA in June 1992. The whole scene consists of 200 bands (0.4–2.5 μm in wavelength range) with a spatial resolution of 20 m/pixel (mpp). Sixteen different land-cover classes are provided in the ground reference map, as shown in Fig. 2(b). The University of Houston data set was taken from an airborne sensor over the area of University of Houston campus and neighboring area, consisting of 349 × 1905 pixels with 144 spectral bands (0.38–1.05 μm). Fifteen different land-cover classes are provided in the ground reference map, as shown in Fig. 3(b).

In the experiments, the average overall accuracy (OA), average accuracy (AA), kappa coefficient (κ), and the individual class accuracy are presented for quantitatively evaluating the proposed method after 10 runs with randomly selected 10, 20, 30, and 40 training samples from each class. Specifically, for the Houston data set, the training samples were randomly selected only from the training set.

A. Parameter Analysis

In this section, the proposed CCAPS-GAN method has four critical parameters (i.e., the sizes ($s \times s$) of local window, the sizes ($k \times k$) of convolutional kernel, the number (M) of feature maps, and the parameter (λ_2) of structured sparse coefficient) that are required to tune. Experiments of parameter analysis on both data sets (i.e., Indian Pines and University of Houston) are performed with ten training samples per class. Tables I–III show the OA values of CCAPS-GAN under different parameters. First, from Table I, it can be seen that CCAPS-GAN can generate satisfactory performance when s is fixed to 27 for Indian Pines. For the University of Houston data set, 9 is a reasonable value for s. Correspondingly, the optimal k is set as {5, 5} and {3, 3} at each convolutional layer. The main reason leading to the difference may be that the former represents a vegetation scenario containing large spatial homogeneity while the University of Houston data set is obtained from an urban area with small homogeneous regions. Furthermore, Table II shows that the optimal combinations of the number of feature maps are {16, 32} and {32, 64}, respectively, for Indian Pines and University of Houston. Finally, Table III lists the OAs of both data sets by using

TABLE IV
CLASSIFICATION WITH SPATIAL–SPECTRAL FEATURES ON THE INDIAN PINES DATA SET (TEN SAMPLES FOR EACH CLASS)

Classifier	SVM	CNN-ELM	Bi-CLSTM	3-D-CNN	AC-GAN	CCAPS-GAN
OA(%)	69.71±1.17	65.37±2.33	61.91±2.24	70.78±2.89	68.04±2.99	**72.31**±1.05
AA(%)	79.16±1.47	78.08±1.09	75.52±2.44	81.40±1.56	80.33±1.38	**81.67**±1.17
$\kappa \times 100$	65.83±3.57	61.13±2.50	57.33±2.83	67.17±3.20	64.04±3.37	**68.74**±1.74
1	95.85±4.17	98.31±1.20	**100.00**±0.00	98.61±2.78	97.92±2.26	98.61±1.53
2	54.97±7.01	38.93±4.16	37.41±13.14	55.36±9.31	**55.13**±8.60	53.13±8.90
3	61.31±5.05	46.68±4.76	39.09±9.66	64.02±13.43	62.04±13.43	**67.08**±6.77
4	78.41±4.45	72.80±3.98	**81.20**±2.73	75.99±11.06	74.78±5.10	71.14±11.67
5	68.50±6.31	**72.67**±7.07	66.67±10.22	72.67±7.61	70.30±8.91	68.09±2.26
6	71.04±2.29	**82.67**±2.54	74.77±3.87	68.58±3.87	70.03±4.31	68.23±7.26
7	100.00±0.00	100.00±0.00	99.02±0.87	100.00±0.00	100.00±0.00	100.00±0.00
8	94.44±4.95	**100.00**±0.00	97.79±1.86	95.73±6.08	93.48±3.58	95.58±4.01
9	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
10	64.79±2.75	55.46±5.79	53.40±11.85	**70.11**±2.51	67.62±4.50	69.31±3.39
11	63.18±5.43	62.15±6.07	61.83±5.78	63.45±2.02	53.42±8.94	**71.13**±2.58
12	75.69±7.14	58.92±6.18	48.43±9.75	70.33±4.76	62.78±8.65	**77.13**±3.63
13	94.49±2.77	**98.59**±0.56	93.16±6.21	96.28±1.75	98.21±1.07	98.03±1.23
14	81.65±5.00	84.00±1.29	83.16±7.17	86.04±5.37	**90.04**±6.12	85.30±3.13
15	90.29±2.85	80.05±4.75	65.98±6.23	86.17±4.32	**90.36**±7.68	84.07±1.70
16	90.66±8.75	96.99±2.48	100.00±0.00	99.10±1.15	100.00±0.00	99.03±1.10

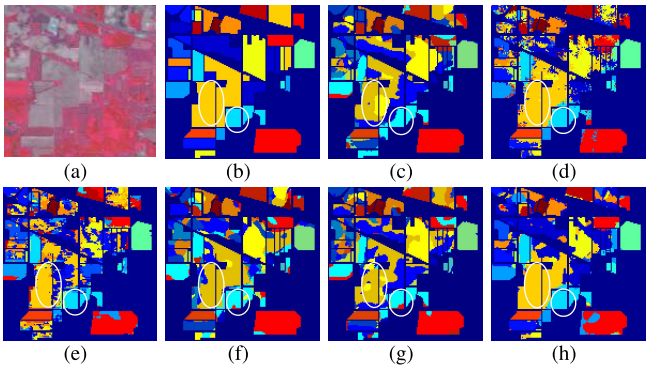

Fig. 2. Comparison of classification results on Indian Pines. (a) Pseudocolor image. (b) Ground-truth map. (c) SVM (69.71%). (d) CNN-ELM (65.37%). (e) Bi-CLSTM (61.91%). (f) 3-D-CNN (70.78%). (g) AC-GAN (68.04%). (h) Proposed (72.31%).

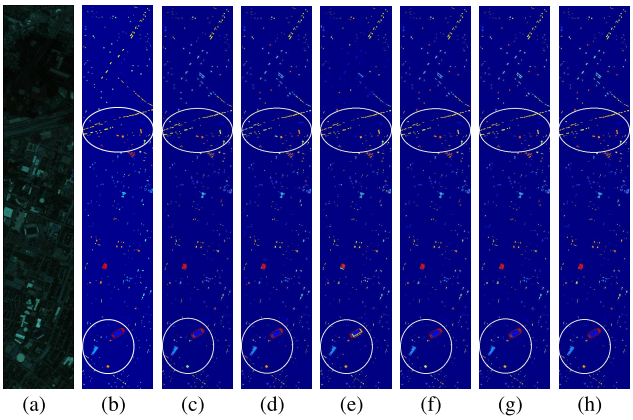

Fig. 3. Comparison of classification results on the University of Houston. (a) Pseudocolor image. (b) Ground-truth map. (c) SVM (62.23%). (d) CNN-ELM (64.20%). (e) Bi-CLSTM (51.52%). (f) 3-D-CNN (67.60%). (g) AC-GAN (68.43%). (h) Proposed (70.69%).

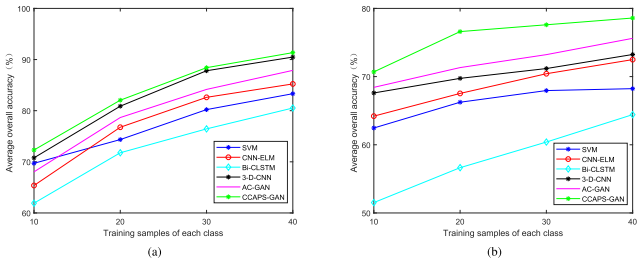

Fig. 4. Average OA of all compared methods under different training samples. (a) Indian Pines data set. (b) University of Houston data set.

CCAPS-GAN with different λ_2 values, which included the ablation experiment (i.e., $\lambda_2 = 0$). Obviously, the highest OAs are achieved by setting $\lambda_2 = 0.001$ for both data sets. It is worth noting that the $L_{2,1}$ regularization will restrict the classification accuracy when λ_2 is large enough.

B. Experimental Results and Analysis

The proposed method is compared with five state-of-the-art HSI processing and classification methods, i.e., SVM [5], CNN-ELM [9], bidirectional-convolutional long short-term memory (Bi-CLSTM) [17], 3-D-CNN [8], and AC-GAN [13]. The SVM classifier with an radial basis function (RBF) kernel is used as the baseline method. The 3-D-CNN method is currently popular for HSI feature extraction and classification. Thus, it is essential to exhibit the performance comparison with 3-D-CNN. In addition, the comparison between the proposed method and the AC-GAN method can verify whether the CCAPS-GAN model can improve the classification performance of the original GAN.

Table IV lists the classification accuracies of different methods on the Indian Pines data set. Obviously, the proposed framework achieves the best classification performance among all considered methods in terms of OA, AA, and κ. Specifically, for the small training samples, the proposed method performs better than the original AC-GAN. The main reason is that the proposed method considered spatial correlation and spectral continuity of HSIs, which makes the generated samples more similar to the real data and efficiently augments the classification network. Moreover, compared with 3-D-CNN and Bi-CLSTM, the proposed method extends a traditional GAN framework with the designed CapsNet discriminator and the updated GAN objective function, which can control the generation of samples and improve the classification performance with a small number of training samples. Fig. 2 shows the classification results of the Indian Pines data set obtained by all considered methods. It can be observed that our method provides better results than others (such as the region marked by the white ellipses).

For the University of Houston data set, a comparison of the proposed method with other ones is shown in Table V. It is

TABLE V
CLASSIFICATION WITH SPATIAL–SPECTRAL FEATURES ON THE UNIVERSITY OF HOUSTON DATA SET (TEN SAMPLES FOR EACH CLASS)

Classifier	SVM	CNN-ELM	Bi-CLSTM	3-D-CNN	AC-GAN	CCAPS-GAN
OA(%)	62.23±2.67	64.20±1.79	51.52±2.77	67.60±1.88	68.43±2.15	**70.69**±1.41
AA(%)	65.12±2.21	66.96±1.49	51.01±1.45	69.65±1.45	70.11±2.01	**72.26**±1.33
$\kappa \times 100$	59.47±2.86	61.57±1.84	46.67±3.27	65.08±1.89	65.93±2.23	**68.25**±1.75
1	81.20±9.20	79.87±6.22	72.17±10.32	81.86±4.67	81.85±6.28	**82.42**±5.75
2	81.02±10.19	58.83±4.36	70.20±9.73	81.05±7.82	83.93±5.71	**84.21**±4.89
3	92.87±3.61	94.85±5.75	42.17±9.17	95.44±1.89	90.10±1.87	**96.44**±1.11
4	71.40±8.09	91.38±6.12	66.38±4.34	89.96±6.77	90.34±3.24	**91.76**±2.28
5	93.56±4.51	97.44±2.08	84.56±7.52	98.01±1.19	98.11±1.08	**98.77**±0.98
6	79.02±12.04	**80.42**±4.21	58.74±6.71	77.62±13.85	78.32±8.46	79.72±6.47
7	43.75±13.46	57.09±10.51	48.50±11.12	56.71±6.73	**67.72**±12.46	53.64±9.47
8	33.43±14.12	**38.84**±10.20	13.29±9.06	34.47±6.92	19.37±9.93	35.33±6.87
9	59.49±12.51	35.88±14.38	43.62±6.00	45.04±14.32	**65.34**±13.26	46.18±8.58
10	46.81±9.12	42.18±9.96	40.92±5.49	46.23±13.19	**53.67**±14.59	47.59±6.44
11	51.04±2.55	58.73±7.75	39.18±4.94	66.22±9.00	56.07±5.65	**75.81**±3.36
12	31.99±8.62	54.95±4.32	44.38±8.41	53.21±7.61	48.13±3.56	**71.95**±4.56
13	24.91±3.66	30.53±5.20	**43.85**±9.69	30.52±7.99	32.63±5.51	32.28±5.78
14	93.12±4.62	92.31±2.81	53.03±4.59	**94.33**±3.59	91.90±3.62	93.52±3.22
15	93.23±1.31	91.12±1.70	44.18±2.20	94.08±1.41	**94.50**±2.03	94.29±1.29

obvious that the proposed framework significantly outperforms other considered methods in terms of OA, AA, and κ. Fig. 3 shows the visual classification results of the University of Houston. As illustrated, CCAPS-GAN exhibits a more smooth map, especially in the area marked by the white ellipses. Despite the fact that the performance of all methods was obtained under few labeled training samples, the proposed method yields less standard deviation in terms of OA, AA, and kappa coefficient in Tables IV and V. As such, by using the structured sparsity, the proposed model resolves the mode collapse problem of GAN and achieves more stable results.

As shown in Fig. 4, the classification performance of all the considered methods with different numbers of training samples is further investigated. It can be seen from both figures that the proposed CCAPS-GAN performs the best for all the cases of 10–40 training samples per class, indicating that it can guarantee the quality of the generated samples for data augmentation to improve the classification performance.

V. CONCLUSION

In this letter, a novel GAN-based deep learning framework has been proposed for HSI classification that can control generated samples for data augmentation to improve the HSI classification performance with limited real training samples. On the one hand, taking advantage of structured sparse $L_{2,1}$ constraint can impose a control on the mode of data being generated and achieve more stable training as well. On the other hand, by utilizing CapsNet and ConvLSTM, the proposed CCAPS-GAN framework can extract low-level features and combine them together with local space sequence information to form powerful high-level contextual features. The proposed method was evaluated on two well-known HSI data sets and compared with SVM, and several the state-of-the-art deep learning-based HSI classification methods. Experimental results have verified that CCAPS-GAN performs better on classification accuracy, and the model is more stable with few training samples.

REFERENCES

[1] P. Duan, X. Kang, S. Li, and P. Ghamisi, "Noise-robust hyperspectral image classification via multi-scale total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1948–1962, Jun. 2019.

[2] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[3] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[4] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, pp. 4085–4098, Nov. 2010.

[5] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[6] H. Yu, L. Gao, W. Liao, B. Zhang, A. Pizurica, and W. Philips, "Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2142–2146, Nov. 2017.

[7] P. Duan, X. Kang, S. Li, P. Ghamisi, and J. A. Benediktsson, "Fusion of multiple edge-preserving operations for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10336–10349, Dec. 2019.

[8] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[9] J. Li, X. Zhao, Y. Li, Q. Du, B. Xi, and J. Hu, "Classification of hyperspectral imagery using a new fully convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 292–296, Feb. 2018.

[10] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019.

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

[12] Z. He, H. Liu, Y. Wang, and J. Hu, "Generative adversarial networks-based semi-supervised learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 10, pp. 1042–1069, Oct. 2017.

[13] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.

[14] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," 2016, *arXiv:1610.09585*. [Online]. Available: http://arxiv.org/abs/1610.09585

[15] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. NISP*, Long Beach, CA, USA, 2017, pp. 3859–3869.

[16] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. NISP*, Montreal, QC, Canada, 2015, pp. 802–810.

[17] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, pp. 17–28, Dec. 2017.