

Please cite the Published Version

Shi, Yue, Han, Liangxiu , Han, Lianghao , Chang, Sheng, Hu, Tongle and Dancey, Darren (2022) A Latent Encoder Coupled Generative Adversarial Network (LE-GAN) for Efficient Hyper-spectral Image Super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 60. p. 5534819. ISSN 0196-2892

DOI: https://doi.org/10.1109/tgrs.2022.3193441

Publisher: Institute of Electrical and Electronics Engineers

Version: Accepted Version

Downloaded from: https://e-space.mmu.ac.uk/630156/

Usage rights: O In Copyright

Additional Information: © 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

A Latent Encoder Coupled Generative Adversarial Network (LE-GAN) for Efficient Hyperspectral Image Super-resolution

Yue Shi, Liangxiu Han*, Lianghao Han, Sheng Chang, Tongle Hu, Darren Dancey

Abstract—Realistic hyperspectral image (HSI) super-resolution (SR) techniques aim to generate a high-resolution (HR) HSI 2 with higher spectral and spatial fidelity from its low-resolution 3 (LR) counterpart. The generative adversarial network (GAN) 4 has proven to be an effective deep learning framework for 5 image super-resolution. However, the optimisation process of 6 existing GAN-based models frequently suffers from the problem 7 of mode collapse, leading to the limited capacity of spectral-8 spatial invariant reconstruction. This may cause the spectral-9 10 spatial distortion on the generated HSI, especially with a large upscaling factor. To alleviate the problem of mode collapse, this 11 work has proposed a novel GAN model coupled with a latent 12 encoder (LE-GAN), which can map the generated spectral-spatial 13 features from the image space to the latent space and produce 14 a coupling component to regularise the generated samples. 15 Essentially, we treat an HSI as a high-dimensional manifold 16 embedded in a latent space. Thus, the optimisation of GAN 17 models is converted to the problem of learning the distributions 18 19 of high-resolution HSI samples in the latent space, making the distributions of the generated super-resolution HSIs closer to 20 those of their original high-resolution counterparts. We have 21 conducted experimental evaluations on the model performance of 22 super-resolution and its capability in alleviating mode collapse. 23 The proposed approach has been tested and validated based on 24 two real HSI datasets with different sensors (i.e. AVIRIS and 25 UHD-185) for various upscaling factors (i.e. $\times 2$, $\times 4$, $\times 8$) and 26 added noise levels (i.e. ∞ db, 40 db, 80 db), and compared with 27 28 the state-of-the-art super-resolution models (i.e. HyCoNet, LTTR, BAGAN, SR- GAN, WGAN). Experimental results show that 29 the proposed model outperforms the competitors on the super-30 resolution quality, robustness, and alleviation of mode collapse. 31 The proposed approach is able to capture spectral and spatial 32 details and generate more faithful samples than its competitors. 33 It has also been found that the proposed model is more robust to 34 noise and less sensitive to the upscaling factor and has been 35 36 proven to be effective in improving the convergence of the generator and the spectral-spatial fidelity of the super-resolution 37 HSIs. 38

Index Terms—Hyperspectral image super-resolution, Genera tive adversarial network, Deep learning.

41

I. INTRODUCTION

Yue Shi, Liangxiu Han, Darren Dancey are with Department of Computing and Mathematics, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester M1 5GD, UK.

Lianghao Han is with Department of Computer Science, Brunel University, UB8 3PH, UK

Sheng Chang is with State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China.

Tongle Hu is College of Plant Protection, Hebei Agriculture University, Baoding 070001, China

Corresponding author: L. Han (e-mail: l.han@mmu.ac.uk)

HE hyperspectral image (HSI) has been widely used in 42 extensive earth observation applications because of the 43 rich information in its abundant spectral bands. However, due 44 to the cost and hardware limitations of imaging systems, the 45 spatial resolution of HSI decreases when the numerous spec-46 tral signals are collected simultaneously [1]-[3]. Due to this 47 drawback, the HSI does not always meet the demands for high-48 accurate earth observation tasks. The HSI super-resolution 49 aiming to estimate a high-resolution (HR) image from a single 50 low-resolution (LR) counterpart is one of promising solutions. 51 Currently, there are mainly two different approaches for HSI 52 super-resolution: 1) the HSI fusion with the HR auxiliary 53 image (e.g. panchromatic image) and 2) the single HSI super-54 resolution without any auxiliary information. Generally, the 55 image fusion approach implements the super-resolution using 56 filter-based approaches through integrating the high-frequency 57 details of HR auxiliary image into the target LR HSI [4], [5], 58 such as component substitution [6], [7], spectral unmixing 59 [8], [9], and Bayesian probability [10], [11]. However, this 60 method highly relies on the high-quality auxiliary image with 61 high imaging cost, which limits its practical applications. In 62 contrast, single HSI super-resolution does not need any other 63 prior or auxiliary information, which has greater practical 64 feasibility. 65

In recent years, the single HSI super-resolution technologies have attracted increasing attention in remotely sensed data enhancement [12]. Particularly, Deep Learning (DL)-based single image super-resolution (SISR) methods have achieved significant performance improvement [13]. The first DL-based method for single image super-resolution was proposed by Dong *et al.* [14], named as the super-resolution convolutional neural network (SRCNN). To recover the finer texture details from low-resolution HSIs with large upscaling factors, Ledig *et al.* [15] proposed a super-resolution generative adversarial network (SRGAN) by introducing a generative adversarial network (GAN). After that, various GAN-based deep learning models have been developed and proven to be effective in improving the quality of image super-resolution [16]–[19].

However, existing GAN-based super-resolution approaches 80 mainly focused on RGB images, in which the reflectance 81 radiance characteristics between the neighbouring spectral 82 channels were not considered in the model training processes. 83 Therefore, using these models for HSI super-resolution di-84 rectly will lead to the absence of spectral-spatial details in 85 the generated images. For example, Fig.1 shows a compar-86 ison between an original high-resolution HSI and its super-87

66

67

68

69

70

71

72

73

74

75

76

77

78

resolution HSI counterpart generated from the SRGAN model 88 [13]. Obvious spectral-spatial distortions can be observed on 89 the generated super-resolution HSI (see the red and yellow 90 frames in Fig.1). Mathematically, recovering spectral-spatial 91 details in super-resolution HSI is an under-determined in-92 verse problem in which a large number of plausible details 93 in the high-resolution image need to be characterised from 94 low-resolution information. The complexity of this under-95 determined issue will exponentially increase with the increased upscaling factor. With high upscaling factors (e.g. higher than 97 $8 \times$), the spectral-spatial details of generated super-resolution 98 HSIs could be distorted. 99



Fig. 1. A comparison between a raw high-resolution (right) HSI and its (8 \times) super-resolution HSI (left) counterpart generated by the SRGAN model [15]. The red frames show the spectral distortion occurs in the learning process, and the yellow frames reveal the loss of spatial details in the super-resolution HSI.

The potential reason behind the spectral-spatial distortion 100 is due to mode collapse in the optimisation process of GANs 101 [20], [21], in which GAN models get stuck in a local minimum 102 and only learn limited modes of data distributions. Some 103 studies have attempted to address mode collapse in GAN 104 models. For instance, Hou et al. [22] improved the diversity 105 of the generator in GAN models and attempted to avoid the 106 mode collapse by adding a reverse generating module and an 107 adaptive domain distance measurement module into the GAN 108 framework. Their findings illustrated that these approaches 109 facilitated solving the insufficient diversity of GAN models 110 in remote sensing image super-resolution. Ma et al. [23] 111 introduced a memory mechanism into GAN models to save 112 feedforward features and extract local dense features between 113 convolutional layers, which showed some effectiveness in 114 increasing spatial details during the reconstruction procedure. 115

To benefit the remarkable super-resolution performance 116 from GAN-based models and address the spectral-spatial dis-117 tortions in HSI super-resolution, in this study, we proposed 118 a novel latent encoder coupled GAN architecture. We treated 119 an HSI as a high-dimensional manifold embedded in a higher 120 dimensional ambient latent space. The optimisation of GAN 121 models was converted to a problem of learning the feature dis-122 tributions of high-resolution HSIs in the latent space, making 123 the spectral-spatial feature distributions of generated super-124 resolution HSIs close to those of their original high-resolution 125 counterparts. Our contributions included: 126

1) A novel GAN-based framework has been proposed to 127 improve HSI super-resolution quality. The improvement was 128 achieved from two aspects. Firstly, for improving the spectral-129 spatial fidelity, a short-term spectral-spatial relationship win-130 dow (STSSRW) mechanism has been introduced to the gener-131 ator in order to facilitate spectral-spatial consistency between 132 the generated super-resolution and real high-resolution HSIs 133 in the training process. Secondly, for alleviating the spectral-134 spatial distortion, a latent encoder has been introduced into the 135 GAN framework as an extra module to make the generator do 136 a better estimation on local spectral-spatial invariance in the 137 latent space. 138

2) A spectral-spatial realistic perceptual (SSRP) loss has 139 been proposed to guide the optimisation of the under-140 determined inverse problem and alleviate spectral-spatial mode 141 collapse issues occurred in the HSI super-resolution process, 142 and benefit on retrieving high-quality spectral-spatial details 143 in the super-resolution HSI, especially for high upscaling 144 factors (e.g. $8\times$). The loss function, SSRP, was able to enforce 145 spectral-spatial invariance in the end-to-end learning process 146 and made the generated super-resolution features closer to the 147 manifold neighbourhood of the targeted high-resolution HSI. 148

The rest of this work is organised as follows: Section 2 149 introduces related works on existing GANs-based methods 150 for HSI super resolution tasks; Section 3 details the proposed 151 approach; Section 4 presents experimental evaluation results; 152 Section 5 concludes the work. 153

II. RELATED WORK

A traditional GAN-based super-resolution model contains two neural networks, a generator producing sample images from low-resolution images and a discriminator distinguishing real and generated images [24]. The generator and discriminator are trained in an adversarial fashion to reach a Nash equilibrium in which the generated super-resolution samples become indistinguishable from real high-resolution samples. 157

Focusing on the spectral and spatial characteristics of HSI 162 data, various adversarial strategies were proposed to improve 163 the GAN performance on HSI super-resolution tasks [25]. For 164 example, Zhu et al [26] proposed a 3D-GAN to improve the 165 generalisation capability of the discriminator in spectral and 166 spatial feature classification with limited ground truth HSI 167 data. Jiang et al [27] designed a spectral and spatial block 168 inserted before the GAN generator in order to extract high-169 frequency spectral-spatial details for reconstructing super-170 resolution HSI data. 171

Some methods for improving the overall visual quality of 172 generated HSIs were also proposed through constructing a 173 reliable mapping function between LR and HR HSI pairs. 174 For example, Li et al [28] proposed a GAN-based model 175 for multi-temporal HSI data enhancement. In their model, 176 a 3DCNN based upscaling block was used to collect more 177 texture information in the upscaling process. Huang et al 178 [29] integrated the residual learning based gradient features 179 between an LR and HR HSI pair with a mapping function in 180 the GAN model, and achieved the HSI super-resolution with 181 an improved spectral and spatial fidelity. 182

The performance of a GAN-based model mainly depends on its generator, its discriminator and loss functions. Therefore, existing studies in improving GAN-based models for HSI resolution focused on their design and optimisation.

187 A. Design of the generator and the discriminator

In the generator where LR data are upscaled to a desired 188 size, upscaling filters are the most important components that 189 influence the performance of the generator in term of accuracy 190 and speed [30], [31]. Ledig et al. [15] employed a deep 191 ResNet with a skip-connection in the generator to produce 192 super-resolution images with $\times 4$ upscaling factor. Jiang *et al.* 193 [16] proposed an edge-enhancement GAN generator in which 194 a group of dense layers were introduced into the generator 195 in order to capture intermediate high-frequency features and 196 recover the high-frequency edge details of HSI data. 197

In regard to the discriminator, it was found that a deeper net-198 work architecture had greater potential in discriminating real 199 images from generated ones [15], [32]. For example, Rangneka 200 et al. [33] proposed a GAN-based deep convolutional neural 201 network with seven convolutional layers in the discriminator 202 for aerial spectral super-resolution. Arun et al. [34] used six 203 3D convolutional filters and three deconvolution filters in the 204 discriminator to discriminate the spectral-spatial features of 205 real HR HSIs from the generated counterparts. 206

In the design of the generator and the discriminator, the 207 computational cost needs to be considered. The upscaling 208 process in the generator can significantly increase the compu-209 tational cost at the scale of n^2 times for an the upscaling factor 210 of n. Meanwhile, the deep learning-based discriminator always 211 requires a large amount of computational time and memory for 212 extracting and discriminating the high-dimensional non-linear 213 mapping features of input data. More efficient generator and 214 discriminator are required for fast and accurate HSI super-215 resolution. 216

217 B. Design of loss functions

The loss function plays a very important role in optimising 218 the performance of GAN models [35], [36]. In the traditional 219 GAN model, the generator and the discriminator are trained 220 simultaneously to find a Nash equilibrium to a two-player 221 non-cooperative game. A min-max loss function is used, it 222 is equivalent to minimising Jensen-Shannon (JS) divergence 223 between the distributions of generative data and real samples 224 when the discriminator is optimal. However, the GAN training 225 is hard, and can be slow and unstable. There are some issues 226 in the original GAN model, such as hard to achieve Nash 227 Equilibrium, the problem of low dimensional supports of 228 sample distributions and mode collapse [37], [38]. To facilitate 229 the training stability and address mode collapse problems in 230 the original GAN model, several improved adversarial loss 231 functions were developed, which can be divided into three 232 categories: 1) the pixel-wised loss, 2) the perceptual loss, and 233 3) the probabilistic latent space loss. 234

In the first category, the pixel-wised mean squared error (MSE) loss is commonly used for measuring the discriminative difference between real and generated data in GAN models

[15]. However, the MSE has some issues, such as the loss of 238 high-frequency details, the over-smoothing problem, and the 239 sparsity of weights [39]-[41]. Some studies have attempted to 240 solve these issues. Chen et al. [42] introduced a sparse MSE 241 function into the GAN model in order to measure the high-242 frequency information in the spatial attention maps of images, 243 their results showed that the GAN with the sparse MSE loss 244 was able to provide more viable segmentation annotations 245 of images. Zhang et al. [43] emphasised that the MSE-loss 246 always led to an over-smoothing issue in the GAN optimisa-247 tion process. Therefore, they introduced a supervised identity-248 based loss functions to measure the semantic differences 249 between pixels in the GAN model. Lei et al. [44] attempted 250 to solve the issue of sparsity of the pixel-wised weights in the 251 GAN model, and proposed two additional metrics, the edge-252 wise KL-divergence and the mismatch rate, for measuring 253 the sparsity of pixel-wised weights and the wide-value-range 254 property of edge weights. 255

In the second category, existing studies used different 256 perceptual losses to balance the perceptual similarity, based 257 on high-level features and pixel similarity. Cha et al. [45] 258 proposed three perceptual loss functions in order to enforce the 259 perceptual similarity between real and generated images, these 260 functions achieved improved performance on generating high-261 resolution image with GAN models. Luo et al. [46] introduced 262 a novel perceptual loss into the GAN based SR model, named 263 as Bi-branch GANs with soft-thresholding (Bi-GANs-ST), to 264 improve the objective performance. Blau et al. [47] proposed 265 a perceptual-distortion loss function in which the generative 266 and perceptual quality of GAN models were jointly quantified. 267 Rad et al. [48] proposed a pixel-wise segmentation annotation 268 to optimise the perceptual loss in a more objective way. 269 Their model achieved a great performance in finding targeted 270 perceptual features. 271

In the third category, Bojanowski et al. [49] investigated the 272 effectiveness of the latent optimisation for GAN models, and 273 proposed a Generative Latent Optimisation (GLO) strategy for 274 mapping the learnable noise vector to the generated images 275 by minimising a simple reconstruction loss. Compared to 276 a classical GAN, the GLO obtained competitive results but 277 without using the adversarial optimisation scheme which was 278 sensitive to random initialisation, model architectures, and the 279 choice of hyper-parameter settings. Training a stable GAN 280 model is challenging. Therefore, Wasserstein GAN [50] was 28 proposed to improve the stability of learning and reduce mode 282 collapse. The WGAN replaced the discriminator model with 283 a critic which scored the realness of a given image in the 284 probabilistic latent space and was trained using Wasserstein 285 loss. Rather than discriminating between real and generated 286 images (i.e. the probability of a generated image being real), 287 the critic maximises the difference between its prediction for 288 real images and generated images (i.e. predict a "realness" 289 score of a generative image). Gulrajani et al. [51] further 290 improved the WGAN training by adding a regularisation term 291 penalising the deviation of the critic's gradient norm with 292 regard to the input, and the model was named as (WGAN-293 GP). 294



Fig. 2. The architecture of the proposed model: the output from the encoder is used to regularise the loss function

III. THE PROPOSED LE-GAN FOR SINGLE HSI SUPER-RESOLUTION

To address the challenge of spectral-spatial distortions 297 caused by mode collapse during the optimisation process, we 298 proposed a novel GAN model coupled with a latent encoder, 299 named as LE-GAN. In the proposed framework, the optimised 300 generator and discriminator were designed to improve the 301 super-resolution performance and reduce the computational 302 complexity. Inspired by the encoder coupled GAN [52], [53], 303 we developed a latent encoder embedded into our GAN 304 framework to facilitate the generator to achieve a better 305 approximation on feature maps, in order to generate ideal 306 super-resolution results. In addition, we designed a spectral-307 spatial realistic perceptual (SSRP) loss function in order to 308 optimise the under-determined inverse problem by providing a 309 trade-off between aligning the distributions of generated super-310 resolution and targeted high-resolution HSIs and increasing the 311 spectral-spatial consistency between them. 312

313 A. Model architecture

295

296

We have made two major changes to the traditional GAN framework: 1) proposed an improved generator, denoted as G, with a simplified ResNet structure, and 2) introduced a latent encoder, denoted as L_E , into the GAN framework. The network architecture is shown in Fig.2. It consists of a generator, a discriminator and an encoder.

1) The architecture of the generator model G: To improve 320 the spectral-spatial reconstruction quality with low distor-321 tion and reduce the computational complexity, a short-term 322 spectral-spatial relationship window (STSSRW) derived gener-323 ator was proposed, denoted as G in our GAN framework. The 324 architecture of the proposed generator G is shown in Fig.3. It 325 serves three functions: low-resolution spectral-spatial feature 326 extraction, STSSRW, and super-resolution HSI reconstruction. 327

Firstly, for the low-resolution spectral-spatial feature extraction, a 3D convolutional filter is introduced. Unlike traditional RGB image super-resolution approaches that use 2D convolutional filters for spatial feature extraction, the HSI superresolution requires processing continuous spectral channels and capturing spectral-spatial joint features from a data cube. Therefore, a 3D convolutional filter is a better choice for modelling both the spectral correlation characteristics and 335 spatial non-local self-similarity. In this study, the convolutional 336 kernel is set to $16 \times 16 \times b$ for a b band HSI input. Nah et 337 al. [54] found that the batch normalisation (BN) layer would 338 normalise the features and get rid of the range flexibility for 339 upscaling features. Therefore, no BN layer is used here to 340 avoid blurring the spectral-spatial information hidden in the 341 convolutional features. 342

Secondly, an STSSRW block is designed to exploit the hierarchical spatial-spectral correlation-state denoted as, H, and further to create the local-global features, F, with low spectral-spatial distortion. It aims to learn the local-global relationship between spectral bands in order to selectively enhance informative band features and spectral-spatial diversity, and achieve low image distortion through modelling the inter-dependencies between high-level features.

More specifically, as shown in Fig. 4a, the low-resolution spectral-spatial data will first be sliced with a sliding window, and then fed into the block. For each data slice B^w , a shortterm correlation-state H^{w-1} is introduced through the feedback connect to correct the local-global relationship between the current data slice w and the previous data slice w-1. Differing from the existing residual connection models that treat the HSI as a whole data cube, the proposed STSSRW approach further divides the data slice B^w into G chunks based on their spectral similarity, that is, $B^w = [BG_1^w, BG_2^w, ..., BG_G^w]$. Then, the short-term correlation-state H^{w-1} in the SRSSRW block are concatenated with each chunk, $B_a^w(g = 1, G)$, with a 1D convolutional operator to update the local-global correlation between the chunks, and a local feature extractor consisting of two ResBlocks (see Fig. 4b) is used to extract the local spectral characteristics in each chunk. The local feature map of the g-th chunk, FG_q^w , can be calculated by:

$$FG_g^w = f_{Res}(f_{Res}(f_{1DConv}(BG_g^w, H^{w-1})))$$
 (1)

After that, the local feature maps $FG_{g}^{w}(g = 1, G)$ from 351 all of the chunks are concatenated and reconstructed as the 352 global spectral features with an upscaling deconvolutional 353 operator, and then another high-level feature extractor with 354 two ResBlocks is employed to create the local-global features, 355 F^w . In the end, a correlation-state update operator including 356 two 1D convolutional filters and one sigmoid activation layer 357 is used to aggregate the band-wise information and update the 358 correlation-state, H^w , for the next iteration. Here, F^w and H^w 359 can be computed with: 360

$$F^{w} = f_{Res}(f_{Res}(f_{Deconv}([FG_{1}^{w}, FG_{2}^{w}, ..., FG_{G}^{w}])))$$
(2)

$$H^{w} = f_{Sigmoid}(f_{1DConv}(f_{1DConv}(f_{1DConv}(f_{Deconv}([FG_{1}^{w}, FG_{2}^{w}, ..., FG_{G}^{w}]))))$$
(3)

where $f_{Sigmoid}, f_{Res}, f_{Deconv}, f_{1DConv}$ represent the operations of Sigmoid, Residual Blocks, Deconvolution and 1D convolution, respectively. 363

Introducing the STSSRW block into the model can not only reduce the spectral dimensionality and computational complexity like normal residual blocks, but also can explore

343

344

345

346

347

348

349



Fig. 3. The architectures of (a) the generator G, (b) the ResBlock component and (c) the UpscaleBlock component. Note: d,w,s are the kernel depth, the kernel width and the stride for a convolutional layer.



Fig. 4. The architectures of STSSRW block. Note: d,w,s are the kernel depth, the kernel width and the stride for a convolutional layer.

the diversity between adjacent spectral bands within a data slice.

Finally, a skip-connect between the input spectral-spatial 369 feature maps and the output local-global feature maps is 370 conducted to selectively enhance the informative spectral-371 spatial structure and suppress the distortion. The subsequent 372 results are fed into an UpscaleBlock to generate the super-373 resolution spectral-spatial details. As shown in Fig.3b, the 374 UpdscaleBlock is a combination of a 3D-convolutional filter 375 and a shuffle layer, in which convolutional filters with a depth 376 of 32 are used to exploit $k^2 \cdot 32 \cdot b$ features for an upscaling 377 factor k. The shuffle layer is used to arrange all the features 378 corresponding to each sub-pixel position in a pre-determined 379 order and aggregate them into super-resolution areas. The size 380 of each super-resolution area is $k \times k$. After this operation, the 381 final feature maps with a size of $(k^2 \cdot 32 \cdot b) \times (H/k) \times (W/k)$] 382 will be arranged into the super-resolution feature maps with 383 a size of $32 \cdot b \times H \times W$, where H and W are the height 384 and width of the super-resolution HSI, respectively. At last, 385 a deconvolution filter is used to decode the feature maps in 386 each area, yielding the super-resolution HSI with enhanced 387 spectral-spatial fidelity. 388

2) The architecture of the discriminator D: The architecture of the proposed discriminator, D, as shown in Fig. 5a, adopts an architecture similar to that used in [15]. But, there



Fig. 5. The architectures of (a) the discriminator, D, and (b) the Maxpool block. Note: d,w and s denote the kernel depth, the kernel width and the stride of a convolutional layer, respectively.

is no sigmoid layer in our model, because the latent space 392 optimisation requires the raw membership without compres-393 sion. Thus, the proposed D mainly contains one convolutional 394 layer, n Maxpool blocks (n = 8 is chosen, following the 395 architectural guidelines summarized by [15]) and two dense 396 layers. The Maxpool block is a combination of a convolutional 397 layer, a BN layer, and a ReLU layer (see Fig. 5b). In this 398 study, eight Maxpool blocks are used to reduce the image 399 dimensionality each time and extract the high-level features, 400 and the resultant feature maps are input into two dense layers 401 to obtain a membership distribution of the feature maps for 402 real or generated HSIs. 403

3) The architecture of the latent encoder, L_E : The latent encoder, L_E , is developed and introduced to the GAN architecture for preventing mode collapse by mapping the generated spectral-spatial features from the image space to the latent space and produces the latent regularisation components in the learning process.

Mathematically, the spectral-spatial features of HSI data, I, generated by the latent encoder, $L_E(I)$, can be decomposed with the singular value decomposition as:

$$L_E(I) = U \times \gamma \times V^T \tag{4}$$

where, U and V are the left and right singular vectors of $L_E(I)$, respectively, and γ can be expressed as:

$$\gamma = \begin{pmatrix} SSD & 0\\ 0 & 0 \end{pmatrix}$$

with $SSD = diag\{\lambda_1, \lambda_2, ..., \lambda_r\}$, the non-zero part of the 410 diagonal matrix after singular value decomposition of HSI data 411 which represents the spectral-spatial distribution (SSD) of the 412 input HSI in the latent space, in which $(\lambda_1 > \lambda_2 > ... > \lambda_r)$ 413 are no-negative singular values, and r is the rank of the matrix. 414 Ideally, in the latent space, the SSD of the super-resolution 415 HSI should be close to that of real high-resolution HSI. 416 However, when the mode collapse occurs (e.g. the generated 417 spectral-spatial features distribution only matches part of the 418 real spectral-spatial feature distribution in the latent space), 419 SSD will concentrate on the top singular values and the rest 420 singular values would be close to zero. Thus, the rank of SSD 421 of the generated HSI will be lower than that of the real HSI. 422 To alleviate the mode collapse, we introduce an extra regularisation term in the loss function to consider the similarity measure of SSD of the spectral-spatial features between the generated data and real data in the latent space. It is defined as:

$$L_{latent} = \mathbb{E}_{I^{hr} \sim Pr(I^{hr})} \|\Delta SSD\|_2 \tag{5}$$

where, $\|\cdot\|_2$ denotes L2 norm, and the ΔSSD is defined as:

$$\Delta SSD = SVD(L_E(I^{hr})) - SVD(L_E(I^{sr}))$$

= $SVD(L_E(I^{hr})) - SVD(L_E(G_{\theta_G}(I^{lr})))$
= $diag\{\lambda_1^{hr} - \lambda_1^{sr}, ..., \lambda_g^{hr} - \lambda_g^{sr}, ..., \lambda_r^{hr} - \lambda_r^{sr}, 0, ..., 0\}$
(6)

where $L_E(I^{hr})$ and $L_E(I^{sr} = G_{\theta_G}(I^{lr}))$ are the 423 outputs from the latent encoder for the real high res-424 olution image, I^{hr} , and the super-resolution image, I^{sr} , 425 generated from low resolution image, I^{lr} , respectively; 426 SVD denotes the singular value decomposition function; 427 $SVD(I^{hr}) = diag\{\lambda_1^{hr}, \lambda_2^{hr}, ..., \lambda_q^{hr}, \lambda_{q+1}^{hr}, ..., \lambda_r^{hr}, 0, ..., 0\}$ 428 are the singular values of $L_E(I^{hr})$, and $SVD(I^{sr}) =$ 429 $diag\{\lambda_1^{sr}, \lambda_2^{sr}, ..., \lambda_g^{sr}, \lambda_{g+1}^{sr}, ..., \lambda_r^{sr}, 0, ..., 0\}$ are the singular values of $L_E(I^{sr} = G_{\theta_G}(I^{lr}))$. This latent regularisation 430 431 term is used to compensate the difference of singular values 432 between real and generated data in the latent space during the 433 optimisation process, which makes the spectral-spatial feature 434 distribution (i.e. singular values) of the generated HSI more 435 closer to the real feature distribution, and further facilitates 436 the diversity of the generated image covers that of the real 437 high-resolution HIS to prevent the singular value degrading 438 and consequently mode collapse. 439

The architecture of the latent encoder is shown in Fig. 6, 440 which consists of eight convolutional layers with an increasing 441 kernel depth by a factor 2 through different layers from 64 to 442 512. The striding operation is used to transfer the spectral-443 spatial features (low-dimensionality) into the latent features 444 (high-dimensionality) once the kernel depth is doubled. The 445 resultant of 512 feature maps are input into two dense layers 446 so that its outputs match the dimension of the latent space. 447

As shown in Fig. 2, L_E receives signals from the generator, 448 $G(I^{lr})$, and the targeted data, I^{hr} , and outputs their represen-449 tations in the latent space which then are used to generate the 450 latent regularisation term, L_{latent}, for the loss function in the 451 model optimisation as described in the section below. To make 452 sure that the outputs of L_E and the real high-resolution HSI in 453 the latent space have the same dimension, L_E is pre-trained by 454 real HSI data. This pre-processing also speeds up the formal 455 optimisation process. 456

457 B. Model optimisation with spectral-spatial realistic percep-458 tual loss

In this study, we treat a low-resolution image as a lowdimension manifold embedded in the latent space, thus the super-resolution HSI can be generated by the parametrised latent space learnt by the model. Theoretically, the generated super-resolution sample, $G_{\theta_G}(I^{lr})$, from a low-resolution sample, I^{lr} , by the generator will be located in a neighbourhood area of its target, I^{hr} , in the latent space.

Latent encoder LE



Fig. 6. The architecture of the latent encoder L_E . Note: d,w and s denote the kernel depth, the kernel width and the stride of a convolutional layer, respectively.

Previous studies [36], [42], [43] used the difference between $G_{\theta_G}(I^{lr})$ and I^{hr} as the generator loss function, described as:

$$|G_{\theta_G}(I^{lr}) - I^{hr}|_1 \le \epsilon \tag{7}$$

However, there are two drawbacks to use this loss func-466 tion in the HSI super-resolution optimisation process. Firstly, 467 the activated features in the latent space are very sparse. 468 The distance based losses rarely consider the spectral-spatial 469 consistency between $G_{\theta_G}(I^{lr})$ and I^{hr} , which leads to the 470 spectral-spatial distortion in the generated super-resolution 471 HSI results. Secondly, the direct bounding on the difference 472 between $G_{\theta_G}(I^{lr})$ and I^{hr} makes it hard to converge because 473 I^{lr} is usually disturbed by the network impairments or random 474 noise. 475

In order to overcome the aforementioned drawbacks, we have designed a spectral-spatial realistic perceptual (SSRP) loss to comprehensively measure the spectral-spatial consistency between $G_{\theta_G}(I^{lr})$ and I^{hr} in the latent space. The formula of the SSRP loss is defined as the weighted sum of the spectral contextual loss, the spatial texture loss, the adversarial loss, and a latent regularisation component, and is shown as follows:

$$L_G^{SSRP} = \lambda \cdot L_{spectral} + \eta \cdot L_{spatial} + \sigma \cdot L_{adversarial} + \mu \cdot L_{latent}$$
(8)

where $L_{spectral}$ is the spectral contextual loss, $L_{spatial}$ is the 476 spatial texture loss, $L_{adversarial}$ is the adversarial loss, and 477 L_{latent} is the latent regularisation component. 478

Based on the SSRP loss, the min-max problem in the GAN model can be described as follows:

r

$$\min_{\theta_G} \max_{\theta_D} L(D_{\theta_D}, G_{\theta_G}) = \min_{\theta_G} \max_{\theta_D} (\lambda \cdot L_{spectral}) + \eta \cdot L_{spatial} + \sigma \cdot L_{adversarial} + \mu \cdot L_{latent})$$
(9)

The details of $L_{spectral}$, $L_{spatial}$, $L_{adversarial}$, are provided below; The definition of L_{latent} can be found in Eq. 5.

1) Spectral contextual loss: $L_{spectral}$ is designed to measure the spectral directional similarity between $G_{\theta_G}(I^{lr})$ and I^{hr} in the latent space, which is defined as follows:

$$L_{spectral} = \mathbb{E}_{z} \{ -log(\frac{1}{N} \cdot \sum_{j} max_{i}A_{ij}) \}$$
(10)

$$A_{ij} = \frac{e^{1 - b_{ij}/n_b}}{\sum_k e^{1 - b_{ij}/n_b}}$$
(11)

where n_b is the band number of an HSI, and b_{ij} is the normalized spectral directional difference defined as:

$$b_{ij} = \frac{c_{ij}}{minc_{ij}} \tag{12}$$

where c_{ij} is used to calculate the directional similarity for both the high-level spectral features and the spectral context between the generated HSI, $G_{\theta_G}(I^{lr})$, and the real HSI, I^{lr} , which is defined as:

$$c_{ij} = \frac{(D_{\mu}(G_{\theta_G}(I_{ij}^{lr})) - D_{\mu}(I_{ij}^{hr})) \cdot (G_{\theta_G}(I_{ij}^{lr}) - I_{ij}^{hr})}{\|D_{\mu}(G_{\theta_G}(I_{ij}^{lr})) - D_{\mu}(I_{ij}^{hr})\|_2 \cdot \|G_{\theta_G}(I_{ij}^{lr}) - I_{ij}^{hr}\|_2}$$
(13)

where $D_{\mu}(\cdot)$ denotes the feature maps obtained from the convolutional layer before the first Maxpooling layer of the discriminator, D.

2) Spatial texture loss: In GAN models, if the loss function only measures the spatial resemblance of the generated and targeted samples, it usually leads to the blurry super-resolution results. In this study, we introduce a spatial texture loss $L_{spatial}$ to measure the texture differences between the feature maps of $G_{\theta_G}(I^{lr})$ and I^{hr} in the latent space. In the $L_{spatial}$, the feature maps of $G_{\theta_G}(I^{lr})$ and I^{hr} before activation are used because they contain more sharp details. $L_{spatial}$ is defined as:

$$L_{spatial} = \mathbb{E}_{z} \{ \frac{1}{W \cdot H} \cdot \sum_{i=1}^{W} \sum_{j=1}^{H} \| D_{\phi}(G_{\theta_{G}}(I_{ij}^{lr})) - D_{\phi}(I_{ij}^{hr}) \|_{2} \}$$
(14)

where $D_{\phi}(\cdot)$ denotes the feature maps obtained from the convolutional layer after the last Maxpooling layer of the discriminator D.

3) Adversarial loss: Along with the spectral contextual loss and the spatial texture loss, an adversarial loss is introduced to facilitate the generator G in reconstructing the image in the ambient manifold space, and fooling the discriminator network. Ladversarial is defined based on the Wasserstein distance [50] between the probability distributions of real data, $P_r(I^{hr})$, and the generated data, $P_g(G_{\theta_G}(I^{lr}))$. Theoretically, $L_{adversarial}$ is strong in alleviating the mode collapse during the training process, because the Wasserstein distance evaluating the similarity between $P_r(I^{hr})$ and $P_q(G_{\theta_G}(I^{lr}))$ rely on the whole samples distributions rather than the individual sample. In other words, there is a penalty would be triggered when the $P_q(G_{\theta_G}(I^{lr}))$ only covers a fraction of $P_r(I^{hr})$, which facilitates the diversity of the generated super-resolution HSI. The goal of $L_{adversarial}$ is to minimise the Wasserstein distance, $W_d(P_r, P_q)$, which is defined as:

$$L_{adversarial} = W_d(P_r, P_g) = \frac{1}{K} \sup_{\|f\|_L < K} \mathbb{E}_{I^{hr} \sim Pr(I^{hr})}[f(I^{hr})] - \mathbb{E}_{I^{lr} \sim Pg(I^{lr})}[f(G_{\theta_G}(I^{lr})]]$$
(15)

where f is the K-Lipschitz function. Suppose we have a parametrised family of functions, $\{f_{w_d}\}_{w_d \in W_d}$, that are all

K-Lipschitz for some K, then the $L_{adversarial}$ can be written as:

$$L_D(P_r, P_g) = \max_{w \in W_d} \mathbb{E}_{I^{hr} \sim Pr(I^{hr})} [f_{w_d}(I^{hr})] -\mathbb{E}_{I^{lr} \sim Pg(I^{lr})} [f_{w_d}(G_{\theta_G}(I^{lr}))]$$
(16)

where W_d is chosen such that the Lipschitz constant of f_{w_d} 489 is smaller than a constant, K. If the probability densities of 490 $P_r(I^{hr})$ and $P_q(I^{lr})$ satisfy the Lipschitz continuous condition 491 (LCC) [55], there is a solution f_{w_d} . Thus, the discriminator 492 is trained to learn a K-Lipschitz continuous function to help 493 compute the Wasserstein distance. The LCC is a strong pre-494 requisite for calculating $W_d(P_r, P_q)$. Therefore, the parame-495 ters, w_d , should lie in a W_d -dimensional manifold in order to 496 meet this constraint. 497

4) The latent regularisation component: In our proposed 498 model, G is a ResNet with global Lipschitz continuity. As 499 described in Section III-A-c, we have introduced a latent 500 encoder, L_E , to compensate the singular values of the spectral-501 spatial features of I^{lr} to the desired I^{hr} . In addition to the op-502 timisation process, the Lipschitz Continuity Condition (LCC) 503 is employed to enforce the local spectral-spatial invariances of 504 G, and map the latent manifold space to a more regularised 505 latent space in case of mode collapse, described as: 506

$$|G_{\theta_G}(I^{lr}) - I^{hr}|_1 \le K \times ||L_E(G_{\theta_G}(I^{lr})) - L_E(I^{hr})||_2$$
(17)

Thus, introducing the regularisation term in the latent space 507 into the loss function (i.e. L_{latent} , see Eq. 5) will make the 508 loss be penalised if the singular values of the spectral-spatial 509 features of a generated super-resolution HSI are updated in a 510 particular direction represented by singular values of SSD). In 511 other words, LLC-derived L_E updating is able to prevent the 512 learning process of each layer from becoming sensitive to the 513 limited direction, which mathematically alleviates the mode 514 collapse, in turn stabilising the optimisation process. 515

IV. EXPERIMENTAL EVALUATION

In this section, we evaluate the effect of proposed LE-517 GEN and determined whether it will improve the super-518 resolution quality and facilitate manifold mapping for solving 519 the problem of mode collapse. Wherein, the developed SSRP 520 loss function plays a key role for both of these prospects. A 521 total of three experiments are designed. The first experiment 522 is to evaluate the optimal parameter combination for the SSRP 523 loss in our proposed model, the second experiment is proposed 524 to evaluate the super-resolution quality, and the last experiment 525 is to evaluate the mode collapse in the model training. 526

The proposed model was trained and tested on real HSI 527 datasets coming from different sensors. It was also compared 528 with five state-of-the-art super-resolution models, including 529 the hyperspectral coupled network (HyCoNet) [56], the low 530 tensor-train rank (LTTR) network [57], the band attention 531 GAN (BAGAN) [58], the super resolution generative ad-532 versarial network (SRGAN) [15], and the Wasserstein GAN 533 (WGAN) [50]. Among them, HyCoNet, LTTR and BAGAN 534 are the state-of-the-art models for HSI super-resolution, while 535 SRGAN and WGAN are the most widely used GAN frame-536 works for image super-resolution. In order to fit the HSI into 537

the SRGAN and WGAN models, a band-wise strategy was employed [59].

540 A. HSI data descriptions

In our experiments, two types of datasets obtained from different sensors were used, one from the public AVIRIS archive, the other from the privately measured UHD-185 data of Guyuan Potato Field (GPF).

1) AVIRIS datasets: Two publicly available HSIs from 545 the AVIRIS data archive were chosen, including the HSIs 546 of Indian Pines (IP) data and the Kennedy Space Center 547 (KSC). Each of them contains 224 hyperspectral bands from 548 $400 \sim 2500 nm$. The HSIs in the KSC dataset were collected 549 by the Kennedy Space Center, Florida, on March 23, 1996. 550 The spatial resolution was 18 m. The HSIs of IP covered 551 the crop planting areas with the spatial resolution of 20 m in 552 North-Western Indiana, USA. In this study, to keep the spectral 553 consistency between different datasets, only the wavelength 554 ranges from invisible to near-infrared ($450 \sim 950 nm$) were 555 considered in our experiments. 556

2) UHD-185 dataset: The UHD-185 dataset contained 557 three privately measured HSIs, denoted as GPF-1, GPF-2, 558 and GPF - 3, in Guyuan Potato Field, Hebei, China. Each 559 of the HSIs was collected by the DJI S1000 UAV sys-560 tem (SZ DJI Technology Co Ltd., Gungdong, China) based 561 UHD-185 Imaging spectrometer (Cubert GmbH, Ulm, Baden-562 Württemberg, Germany) in 2019. All the images were obtained 563 at a flight height of 30 m, with 220 bands from visible to 564 near-infrared bands between 450 and 950 nm and a spatial 565 resolution close to 0.25m per pixel. 566

567 B. Evaluation metrics

The evaluation metrics include 1) the metrics for evaluating super-resolution quality and robustness and 2) the metrics for evaluating mode collapse of GANs.

Evaluation metrics for super-resolution quality and ro- bustness assessment: In total, five spectral-spatial evaluation
 metrics were employed for the super-resolution quality assessment.

These five metrics are 1) Information entropy associated 575 peak signal-to-noise (PSNR), 2) Spatial texture associated 576 structural similarity index (SSIM), 3) Perception-distortion 577 associated perceptual index (PI), 4) spectral reality associated 578 spectral angle mapper (SAM) and 5) Spectral consistency 579 associated spectral relative error (SRE). Among them, the 580 PSNR and SSIM were widely used in the evaluation of image 581 quality [60], the larger the score of PSNR or SSIM the higher 582 the image-quality. 583

The PSNR is defined as:

$$PSNR(I^{hr}, I^{sr}) = 10 \cdot \log_{10}(255^2/MSE(I^{hr}, I^{sr}))$$
 (18)

where $MSE(I^{hr}, I^{sr})$ is the mean squared error between the real HR HSI, I^{hr} , and the generated HR HSI through superresolution, I^{sr} . The PSNR goes to infinity as the MSE goes to zero.

The SSIM is defined as:

$$SSIM(I^{hr}, I^{sr}) = l(I^{hr}, I^{sr}) \cdot c(I^{hr}, I^{sr}) \cdot s(I^{hr}, I^{sr})$$
(19)

where $l(I^{hr}, I^{sr})$, $c(I^{hr}, I^{sr})$, and $s(I^{hr}, I^{sr})$ are the difference measures for luminance, contrast, and saturation between real and generated HR HSI pairs, respectively. The details can be found in [61].

However, the numerical scores of PSNR and SSIM are not always correlated well with the subjective image quality. Therefore, Blau *et al.* [62] proposed an index, PI (Perception Index), as a compensatory reference for the image quality evaluation. The lower the PI value is, the higher the perceptual quality of the image. The PI is defined by two non-referenced image quality measurements, MA [63] and NIQE [64], described as:

$$PI(I^{hr}, I^{sr}) = \frac{1}{2}((10 - MA(I^{hr}, I^{sr}))) + NIQE(I^{hr}, I^{sr})$$
(20)

In order to measure the spectral distortion, the spectral angle mapper(SAM), was used to calculate the average angle between a super-resolution HSI and its targeted high-resolution HSI. The SAM is defined as:

$$SAM(I^{hr}, I^{sr}) = \frac{1}{n} \sum \arccos(\frac{I^{hr} \cdot I^{sr}}{\|I^{hr}\|_2 \cdot \|I^{sr}\|_2})$$
(21)

where n is the pixel number of the HSI.

To evaluate the pixel-wised spectral reconstruction quality, the spectral relative error (SRE) was also used as a metric, defined as:

$$SRE(I^{hr}, I^{sr}) = \left[\frac{1}{n_b} \sum_{i=1}^{n_b} \|(I_i^{hr}, I_i^{sr})\|^2\right]^{\frac{1}{2}}$$
(22)

where the n_b is the band number of an HSI.

2) Evaluation metrics for mode collapse of GANs : Three metrics for GANs, Inception Score (IS), Frechet Inception Distance (FID) and non-referenced spectral score (Non-ref Score), are employed to measure the mode collapse through monitoring the mode collapse and spectral-spatial distortion in the model training process [65]-[67]. The IS measures both the image quality of generated HSIs and their diversity, reflecting the probability of mode collapse in the model training process. In GANs, it is desirable for the conditional probability, $p(I^{hr}|G(I^{lr}))$ to be highly predictable (low entropy), that is, the probability density function is less uniform. The diversity of the generated image can be measured with the marginal probability, $p(I^{hr}) = \int p(I^{hr}|G(I^{lr}))dI^{lr}$. The less uniform (low entropy) the marginal probability is, the less the diversity of the generated image is. Through computing the KL-divergence between these two probability distributions, the IS is computed with the equation below:

$$IS = exp[\mathbb{E}_{I^{lr} \sim p(I^{lr})}[\mathbb{D}_{KL}(p(I^{hr}|G(I^{lr}))||p(I^{hr}))]] \quad (23)$$

The Frechet Inception Distance (FID) score is a metric calculating the distance between the feature vectors extracted from real and generated images. It was used to evaluate the quality of GAN generated images, and a lower FID value correlates with a higher image quality. More importantly, the FID is sensitive to mode collapse. Through modelling the distribution of the features extracted from an intermediate layer with a multivariate Gaussian distribution, the FID between 592

the real image and generated images is calculated using the following equation,

$$FID = ||M_{hr} - M_{sr}||_2^2 + Tr(C_{hr} + C_{sr} - 2(C_{hr} \times C_{sr})^{1/2})$$
(24)

where M_{hr} and M_{sr} refer to the feature-wise means of the real high-resolution HSI and the generated super-resolution HSI in discriminator model, respectively, and C_{hr} and C_{sr} are the covariance matrix for the real and generated feature vectors, respectively.

The non-referenced spectral score (Non-ref Score), which 599 is firstly proposed by [67], is a no-reference assessment 600 method for hyperspectral distortion. It was used to exploit the 601 spectral distortion of GAN generated images. A lower score 602 correlated with a lower reconstructed spectral distortion. The 603 Non-ref Score is calculated by averaging the distances over all 604 intermediate features. The detailed information about Non-ref 605 Score can be referred to [67]. 606

607 C. Experimental configuration

In our experiments, the raw HSIs were labelled as HR 608 samples. The LR samples were generated by down-sampling 609 the HR samples with three scaling factors, $\times 2$, $\times 4$ and $\times 8$, 610 based on the bi-cubic interpolation approach [68]. For the 611 AVIRIS datasets, the KSC data was used for the model training 612 and test, and the IP data was used for the independent test. 613 For the UHD-185 dataset, the GPF - 1 and GPF - 2 were 614 used for training and test, and the GPF - 3 was used for 615 the independent test. More specifically, for the training/test 616 datasets, the HR HSI was cropped into a series of sub-images 617 with a spatial size of 384×384 , and the corresponding LR 618 data was respectively cropped to 192×192 , 96×96 , and 619 48×48 . After this operation, a total of 896 HR and LR HSI 620 pairs were generated from the AVIRIS dataset, and 952 HR 621 and LR HSI pairs were generated from the UHD-185 dataset, 622 in which 70% of image pairs were randomly selected as the 623 training set and the rest 30% of image pairs were used as the 624 test set. 625

The training process was divided into two stages. In the first 626 stage, the discriminator D and the latent encoder L_E were pre-627 trained over 5,000 iterations on the raw HR HSI dataset to get 628 initial weights. The Adam optimiser was used by setting the 629 forgetting factors, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, a small scalar 630 constant $\epsilon = 10^{-7}$ and the learning rate $= 10^{-4}$ [69]. In the 631 second stage, the discriminator, the generator, and the latent 632 encoder were jointly trained for over 10,000 times, until they 633 converged. The Adam optimiser with the same parameters was 634 used. All of the training were performed on NVIDIA 1080Ti 635 GPUs. 636

637 D. Experiment 1: the parameter selection for the SSRP loss function

To achieve an optimal performance, an optimised combination of the parameters in the SSRP loss function Eq.(8), λ , η , σ and μ , needs to be found. In this study, a traversal method was employed to search the optimal parameter combination. These parameters were traversed in the range of 0 to 100 with

TABLE I The top five combinations of the parameters, λ , η , σ and μ for the super-resolution HSI generation with the scaling factors of $\times 2$, $\times 4$ and $\times 8$ based on the values of spectral-spatial evaluation metrics after 10,000 iterations.

Scaling factor	No.	$(\lambda, \eta, \sigma, and \mu)$	PSNR	SSIM	PI	SAM	SRE
	1	(12.8, 12.9, 0.008, 0.015)	31.738	0.982	3.782	5.011	8.383
	2	(12.8, 12.8, 0.009, 0.016)	31.716	0.945	3.884	5.155	8.461
$\times 2$	3	(12.7, 12.8, 0.007, 0.014)	31.712	0.963	3.87	5.115	8.469
	4	(12.8, 12.8, 0.008, 0.014)	31.712	0.943	3.849	5.161	8.482
	5	(12.6, 12.8, 0.006, 0.017)	31.708	0.926	3.876	5.174	8.499
	1	(12.4, 12.4, 0.006, 0.015)	31.417	0.903	3.765	4.942	8.219
	2	(12.4, 12.5, 0.009, 0.014)	31.395	0.901	3.764	5.075	8.267
$\times 4$	3	(12.4, 12.3, 0.007, 0.015)	31.375	0.893	3.765	5.013	8.279
	4	(12.5, 12.8, 0.007, 0.014)	31.359	0.891	3.767	5.017	8.276
	5	(12.5, 12.8, 0.006, 0.017)	31.322	0.898	3.819	5.065	8.331
	1	(12.3, 12.3, 0.005, 0.015)	29.881	0.931	3.672	4.741	8.672
	2	(12.4, 12.3, 0.006, 0.014)	29.851	0.902	3.663	4.828	8.726
×8	3	(12.4, 12.2, 0.004, 0.014)	29.816	0.885	3.583	4.797	8.753
	4	(12.5, 12.5, 0.005, 0.014)	29.828	0.923	3.617	4.866	8.679
	5	(12.4, 12.6, 0.005, 0.015)	29.791	0.885	3.634	4.817	8.733

a fixed step of 0.001 for the range of 0 to 1, and a fixed step 644 of 0.1 for the range of 1 to 100. The selection of parameter 645 combinations was based on the spectral-spatial quality of 646 generated super-resolution HSIs measured with five evaluation 647 metrics, PSNR, SSIM, PI, SAM, and SRE. Table. I lists the 648 top five parameter combinations and the corresponding values 649 of these metrics for generating the super-resolution HSI with 650 the scaling factors of $\times 2$, $\times 4$ and $\times 8$. It can be observed that 651 all the parameters after optimisation are located in a relatively 652 small range, for example, 12.3-12.8 for λ and η , 0.004-0.009653 for σ and 0.014 - 0.017 for μ . In the following experiments, 654 we employed the average values of the best parameters for 655 various scaling factors, thus, $\lambda = 12.5, \eta = 12.5, \sigma = 0.0063$, 656 and $\mu = 0.015$. 657

E. Experiment 2: model robustness and super-resolution quality assessment

To evaluate the robustness and generalizability of the proposed model, we have evaluated our model on both testing datasets and independent datasets.

1) Model assessment on the testing datasets: As described 663 in Section IV-C, we divided the dataset into the training 664 and testing datasets. The performance of the proposed model 665 for hyperspectral super-resolution with three upscaling factors 666 $(\times 2, \times 4 \text{ and } \times 8)$ was evaluated on testing datasets including 667 AVIRIS (KSC) and UHD-185(GPF-1 and GPF-2), and com-668 pared with five state-of-the-art competition models. To assess 669 the model robustness to noise, the model was also evaluated 670 on the datasets with artificially added Gaussian white noise 671 at three different levels (∞ , 40db and 80db) to each of the 672 spectral bands of low-resolution HSIs. To facilitate ranking 673 the models in terms of reconstruction quality, five most widely 674 used evaluation metrics, PSNR, SSIM, PI, SAM, and SRE, 675 were chosen. Specifically, PSNR, SSIM and PI were used 676 to measure the spatial reconstructed quality from the aspects 677 of information entropy, spatial similarity, and perception dis-678 tortion, respectively. The higher PSNR and SSIM scores and 679 the lower PI scores indicate the higher spatial reconstruction 680 quality. In addition, the SAM and SRE scores were used for the 681 spectral distortion measurement from the aspects of spectral 682 angle offset and amplitude difference, respectively. The lower 683

- 658

659

660

661

values of SAM and SRE scores indicate the higher spectral reconstruction quality.

Table II and Table III provide the average scores of PSNR, 686 SSIM, PI, SAM, and SRE of HSI super-resolution results from 687 the proposed model and its five competitors using the AVIRIS 688 and UHD-185 testing datasets, respectively. In general, the 689 results on both datasets consistently show that the proposed 690 LE-GAN model achieves the highest PSNR and SSIM values 691 and the lowest PI, SAM and SRE values for all three dif-692 ferent upscaling factors and three added noise levels (see the 693 highlighted values in Table II and Table III). This means that 694 LE-GAN achieves the best spectral and spatial fidelity and 695 super-resolution quality. 696

A more detailed analysis of the results for the model performance evaluation was performed from two aspects: (1) Super-resolution performance under various upscaling factors (2) Model robustness against different noise levels. Since the results in Table II and Table III have the similar patterns for all the models, here we only present the analyses and assessment using the results on AVIRIS data (i.e. Table II):

(1) Among three upscaling factors, the LE-GAN based 704 super-resolution with the smallest upscaling factor $\times 2$ and 705 without added noise (i.e. ∞ db) achieves the best spectral and 706 spatial reconstruction quality. The best scores of PSNR, SSIM, 707 PI, SAM, and SRE are 35.513, 0.898, 3.052, 4.207, and 8.379, 708 respectively, which are closer to the real high-resolution HSI 709 (i.e. 35.981 for PSNR, 0.912 for SSIM, 3.011 for PI, 4.142 710 for SAM, and 8.019 for SRE), compared to its competitors. 711 The similarities (i.e. the ratio between the super-resolution 712 HSI and the real high-resolution HSI) reach 98.7%, 98.46%, 713 98.66%, 98.45%, and 95.7%, respectively. In addition, for a 714 given added noise level, the spectral and spatial quality of 715 the LE-GAN generated super-resolution HSIs are more stable 716 between the upscaling factors of $\times 2$ and $\times 4$. For example, 717 under the added noise level of 80db, the PSNR, SSIM, PI, 718 SAM and SRE scores are 35.225, 0.835, 3.171, 4.221, and 719 8.839 for $\times 2$ upscaling factor, and increasing the upscaling 720 factor to $\times 4$ only causes the slight changes to these scores 721 which are 34.975, 0.804, 3.364, 4.362 and 9.062, respectively. 722 The consistency ratios (i.e. the ratio between the $\times 2$ and $\times 4$ 723 super-resolution HSI) are 99.29%, 96.29%, 94.26%, 96.78%, 724 and 97.54%, respectively. In contrast, a larger performance 725 degradation occurs on the spectral and spatial reconstruction 726 quality of the competitors. For example, with regard to the 727 WGAN, the second best model in terms of PSNR and SSIM, 728 the scores of PSNR, SSIM, PI, SAM, and SRE under non-729 added noise level are 33.729, 0.826, 3.867, 7.248, 14.152 730 with $\times 2$ upscaling, but change to 30.035, 0.807, 4.476, 7.922, 731 14.361 with $\times 4$ upscaling, showing the performance degrada-732 tions of 10.95%, 2.3%, 13.6%, 8.5%, and 1.5%, respectively. 733 Although the degradations of PSNR, SSIM, PI, SAM, SRE 734 scores can be observed on all the models for $\times 8$ upscaling, 735 the degradation rate of these scores from the proposed LE-736 GAN is the smallest. For example, under the non-added noise 737 $(\infty \text{ db})$, the SNR, SSIM, PI, SAM, SRE scores of LE-GAN 738 based super-resolution HSIs for $\times 8$ upscaling are 32.078, 739 0.784, 3.988, 4.711 and 8.986, respectively, which are 21.03%, 740 6.76%, 31.4%, 15.9% and 25.4% higher than those based on 741

 the second best models (i.e. the WGAN in terms of SSIM
 742

 (26.591) and the BAGAN in terms of PSNR (26.591), PI
 743

 (5.814), SAM (5.602), and SRE(12.048)).
 744

(2) With regard to the model robustness to noise, the 745 proposed LE-GAN shows the best performance on the spectral 746 and spatial reconstruction for a given upscaling factor in 747 comparison with its competitors, although the degradation is 748 observed with increased noise levels. The smaller the upscaling 749 factor is, the more robust the model is. The most robust results 750 against noise are at the upscaling factor of $\times 2$. Only 3.6%, 751 7.9%, 14.67%, 13.11%. and 7.33% degradations of the PSNR, 752 SSIM, PI, SAM, and SRE scores of LE-GAN-based super-753 resolution results occur when the added noise level increases 754 from non-added (∞ db) to 40 db (see Table II). In contrast, 755 the added noise-induced degradations to the results from the 756 WGAN (the second best model for $\times 2$ upscaling factor) 757 are much higher, reaching 20.29%, 8.35%, 26.65%, 20.23%, 758 and 21.61%, respectively. In addition, when the upscaling 759 factor increases from $\times 2$ to $\times 8$, the added noise-induced 760 degradations on the PSNR, SSIM, PI, SAM and SRE scores 761 of the LE-GAN super-resolution results are 9.58%, 11.86%, 762 19.42%, 14.02%, and 11.25%, which are acceptable for the 763 super-resolution with a high upscaling factor and high noises. 764 In contrast, a more serious deterioration can be observed in 765 the results from its competitors. For example, the added noise-766 induced degradations on the PSNR, SSIM, PI, SAM, SRE of 767 the BAGAN-based super-resolution results, the second best 768 model, are 21.72%, 10.57%, 15.6%, 13.45% and 15.05%, 769 respectively, for an upscaling factor of $\times 2$, but change to 770 24.07%, 23.12%, 13.23%, 15.55%, 19.84% for an upscaling 77 factor of $\times 8$. 772

2) Model assessment on the independent test datasets: The 773 proposed model has also been evaluated on two independent 774 test datasets, AVIRIS (IP) and UHD-185 (GPF-3), which 775 were not involved in the model training. Fig. 7 illustrates 776 a comparison of five evaluation metrics (PSNR, SSIM, PI, 777 SAM and SRE) between the proposed model and its five 778 competitors. The average value and standard deviation of 779 each metric were calculated based on the measures at three 780 noise levels, ∞ , 80db and 40db. Compared to its competitors, 781 the proposed model achieves the highest average values and 782 lowest standard deviations for PSNR, SSIM, and the lowest 783 average values and the lowest standard deviation for PI, SAM 784 and SRE, across three upscaling factors on both AVIRIS test 785 dataset (see Fig. 7a) and UHD-185 test dataset (see Fig. 7b). 786 That is, the proposed model achieves the best performance on 787 super-resolution. Similar to the evaluation results in Subsection 788 IV-E1, overall the second best model on the independent test 789 datasets is WGAN for the spatial information reconstruction 790 measure (e.g. PSNR, SSIM), and BAGAN for the spectral 791 information reconstruction measure (e.g. SAM, SRE). 792

It can also be observed that the changes of these metrics are relatively small with the increase of the upscaling factor. When the upscaling factor increases from $\times 2$ to $\times 4$, the average values of SSIM, PI, SAM, and SRE from the proposed model almost stay the same; When the upsampling factor increases from $\times 4$ to $\times 8$, the changes of these metrics are much smaller compared to those from its competitors.

TABLE II A QUANTITATIVE COMPARISON OF HSIS SUPER-RESOLUTION SPECTRAL AND SPATIAL QUALITY IN TERMS OF THE AVERAGE PSNR, SSIM, PI, SAM, SRE SCORES USING THE PROPOSED MODEL AND FIVE COMPETITION MODELS ON TEST DATASETS, AVIRIS (KSC DATA), WITH VARIOUS UPSCALING FACTORS AND ADDED NOISE LEVELS. NOTE THAT THE PSNR/SSIM/PI/SAM/SRE SCORES FOR THE KSC DATA (I.E. REAL HIGH-RESOLUTION HSI) ARE 35.981, 0.912, 3.011, 4.142, AND 8.019 RESPECTIVELY, THE HIGHER PSNR, SSIM AND THE LOWER PI, SAM, SRE, THE BETTER THE SPECTRAL AND SPATIAL FIDELITY.

	SNR(db)		HyCoNet	LTTR	BAGAN	SRGAN	WGAN	LE-GAN
	∞	PSNR	31.213	29.495	32.177	32.421	33.729	35.513
		SSIM	0.792	0.729	0.766	0.809	0.826	0.898
		PI	4.181	4.269	3.672	4.015	3.867	3.052
		SAM	6.491	6.515	5.485	9.014	7.248	4.207
		SRE	10.813	10.145	10.476	15.438	14.152	8.379
	80	PSNR	28.751	24.981	29.121	30.106	32.945	35.225
		SSIM	0.719	0.735	0.761	0.756	0.808	0.835
AVIRIS $(\times 2)$		PI	4.178	4.819	3.766	3.991	4.196	3.171
		SAM	6.622	7.164	5.961	10.961	7.297	4.221
		SRE	12.447	11.913	11.216	19.301	17.914	8.839
	40	PSNR	23.777	23.253	25.187	26.181	26.886	34.223
		SSIM	0.571	0.634	0.685	0.715	0.757	0.827
		PI	4.751	3.182	4.351	5.489	5.272	3.577
		SAM	7.591	7.516	6.338	11.177	9.087	4.842
		SRE	14.574	12.946	12.332	20.028	18.053	9.042
	∞	PSNR	27.216	27.841	29.177	26.105	30.035	35.367
		SSIM	0.774	0.725	0.762	0.799	0.807	0.835
		PI	4.781	4.474	4.291	4.812	4.476	3.061
		SAM	6.514	6.533	5.566	9.729	7.922	4.272
		SRE	12.75	10.811	10.641	16.68	14.361	8.431
	80	PSNR	24.816	25.896	27.048	22.777	28.183	34.975
		SSIM	0.559	0.665	0.513	0.524	0.709	0.804
AVIRIS $(\times 4)$		PI	4.514	5.889	4.356	5.037	5.101	3.364
		SAM	6.571	7.031	5.685	11.582	8.926	4.362
		SRE	13.551	11.562	11.406	17.799	18.321	9.062
	40	PSNR	21.714	22.669	26.363	25.934	29.093	33.041
		SSIM	0.315	0.651	0.479	0.668	0.604	0.786
		PI	6.051	6.072	4.854	6.223	5.912	3.376
		SAM	6.771	7.736	6.268	12.091	10.138	4.443
		SRE	15.041	13.203	13.114	18.158	18.556	9.383
	∞	PSNR	19.871	21.747	26.591	24.589	25.332	32.078
		SSIM	0.622	0.638	0.718	0.668	0.731	0.784
		PI	6.835	6.322	5.814	6.536	5.902	3.988
		SAM	7.69	6.992	5.602	10.179	8.126	4.711
		SRE	14.361	14.302	12.048	16.943	16.092	8.986
	80	PSNR	17.821	20.402	24.073	20.603	24.369	30.291
		SSIM	0.552	0.493	0.692	0.605	0.663	0.775
AVIRIS $(\times 8)$		PI	7.421	7.094	6.411	6.942	6.116	4.326
		SAM	7.72	7.667	6.283	11.499	9.321	4.732
		SRE	16.361	15.534	13.194	18.213	18.101	9.768
	40	PSNR	14.84	18.344	20.191	19.532	21.572	29.003
		SSIM	0.316	0.418	0.552	0.511	0.547	0.691
		PI	7.622	7.954	6.701	7.924	7.366	4.949
		SAM	9.172	9.061	6.616	12.541	8.525	5.479
		SRE	18.219	18.836	15.031	19.597	19.598	10.125

These findings suggest that the proposed model overcomes the drawback associated with spectral-spatial reconstruction under the noises interferences compared to its competitors. Moreover, the proposed model is less sensitive to the upscaling factor, and has a good performance even with a large upscaling factor (e.g. ×8).

3) Visual Analysis of generated super-resolution HSIs with 806 a large upsampling factor ($\times 8$): To demonstrate the per-807 formance improvement of the proposed model in spectral-808 spatial fidelity, visual analyses on generated super-resolution 809 HSI samples have been performed. Fig. 8 displays the results 810 from independent test datasets (IP and GPF-3). Although 811 the visualisation results from the proposed method and its 812 competitors are similar, the image edges from the LE-GAN 813 are sharper than those from the competitors. For example, 814 the internal textures of the bare-soil shown as grey in the 815 false-colour images almost disappear in the generated super-816 resolution IP images by the HyCoNet, LTTR, and BAGAN 817 (the second, third, and forth images in the first row of Fig. 8). 818 These findings suggest that the LE-GAN provides improved 819 spatial quality in general. 820

To further visualise the super-resolution details on spatial



Fig. 7. A comparison of five evaluation metrics (PSNR, SSIM, PI, SAM, and SRE) between the proposed model and its five competitors evaluated on independent test datasets (a) AVIRIS (IP) and (b) UHD-186 (GPF-3). The average values and standard deviations of each metrics are calculated across three different noise levels(∞ , 80*db* and 40*db*).



Fig. 8. A sample of super-resolution results $(\times 8)$ from our models and its five competitors on independent test datasets

TABLE III

A QUANTITATIVE COMPARISON OF HSIS SUPER-RESOLUTION SPECTRAL AND SPATIAL QUALITY IN TERMS OF THE AVERAGE PSNR, SSIM, PI, SAM, SRE SCORES USING THE PROPOSED MODEL AND FIVE COMPETITION MODELS ON TEST DATASETS, UHD-185 (GPF-1 AND GPF-2), WITH VARIOUS UPSCALING FACTORS AND ADDED NOISES. NOTE THAT THE PSNR/SSIM/PI/SAM/SRE SCORES FOR THE KSC DATA (I.E. REAL HIGH-RESOLUTION HSI) ARE 38.915, 0.992, 4.418, 6.942, AND 10.519 RESPECTIVELY. THE HIGHER PSNR AND SSIM AND THE LOWER PI, SAM, AND SRE, THE BETTER THE SPECTRAL AND SPATIAL FIDELITY.

	SNR(db)		HyCoNet	LTTR	BAGAN	SRGAN	WGAN	LE-GAN
	∞	PSNR	33.238	32.689	34.642	36.009	37.697	38.575
		SSIM	0.874	0.875	0.851	0.897	0.879	0.979
		PI	5.11	5.454	4.799	5.185	5.207	4.323
		SAM	9.174	8.124	7.266	11.848	9.904	6.893
		SRE	15.755	14.711	12.677	17.341	15.43	10.295
	80	PSNR	30.484	29.196	31.913	32.28	35.37	37.625
		SSIM	0.796	0.831	0.841	0.846	0.873	0.922
UHD-185 $(\times 2)$		PI	5.911	5.113	5.002	5.341	5.523	4.238
		SAM	9.331	9.073	7.921	13.05	9.849	6.899
		SRE	14.355	13.441	13.039	20.549	19.842	10.711
	40	PSNR	28.834	27.137	28.904	29.296	33.752	36.976
		SSIM	0.593	0.632	0.758	0.776	0.833	0.873
		PI	5.905	6.223	5.636	6.536	6.574	4.592
		SAM	10.53	10.311	10.172	13.597	11.328	6.508
		SRE	16.029	14.603	14.523	21.425	19.695	10.823
	∞	PSNR	29.855	30.341	34.111	33.285	34.863	38.303
		SSIM	0.717	0.804	0.827	0.861	0.834	0.892
		PI	6.161	5.755	5.309	6.194	5.944	4.391
		SAM	9.408	9.187	7.311	12.652	9.937	6.563
		SRE	13.81	12.442	12.208	18.622	15.776	10.005
	80	PSNR	27.046	27.702	30.396	30.731	33.866	37.061
		SSIM	0.661	0.717	0.666	0.813	0.814	0.816
UHD-185 $(\times 4)$		PI	6.791	6.995	5.741	6.061	6.276	4.463
		SAM	9.376	9.909	7.956	13.939	10.09	6.554
		SRE	15.464	14.963	17.128	21.509	19.706	10.361
	40	PSNR	26.948	26.891	27.917	25.844	32.217	36.556
		SSIM	0.459	0.606	0.567	0.751	0.687	0.731
		PI	7.521	7.197	6.136	7.621	7.311	4.751
		SAM	12.679	10.262	9.028	13.99	12.912	6.661
		SRE	16.518	14.396	14.385	22.221	20.411	10.939
	∞	PSNR	22.675	25.375	27.831	27.367	29.012	35.397
		SSIM	0.705	0.688	0.784	0.734	0.821	0.852
		PI	7.908	7.557	6.835	7.863	7.153	5.046
		SAM	9.487	9.668	7.827	13.942	10.251	7.104
		SRE	16.378	16.287	13.64	19.011	17.271	10.145
	80	PSNR	21.879	23.88	26.586	24.029	27.604	33.166
		SSIM	0.677	0.549	0.564	0.679	0.755	0.801
UHD-185 (×8)		PI	8.541	8.258	7.426	8.393	7.611	5.649
		SAM	9.814	10.52	8.931	14.092	12.152	7.669
		SRE	17.627	16.941	15.946	22.224	21.887	11.333
	40	PSNR	17.819	21.166	23.199	22.251	24.132	31.519
		SSIM	0.419	0.474	0.541	0.591	0.632	0.715
		PI	8.674	9.154	7.872	9.058	8.755	6.229
		SAM	13.694	11.175	8.559	15.236	14.212	7.945
		SRE	20.042	20.604	16.103	25.159	22.086	11.162

and spectral fidelity, some representative false-colour compos-822 ite image patches and the spectral curves of the super-resolved 823 HSI patches from independent test dataset (GPF-3) are shown 824 in Fig. 9. It is obvious that the brightness, contrast, and internal 825 structures of the false-colour images generated by the LE-826 GAN are more faithful to real HR data. For example, the land 827 cover textures in the LE-GAN generated image (the second 828 image from the right in the second row of images in Fig. 9) are 829 clearer, compared to the images generated by the competitors 830 (e.g. the HyCoNet and LTTR based images) in which the edges 831 of streets are fuzzy. Moreover, the spectral curves from the LE-832 GAN generated images are more consistent with those from 833 real HR HSI data. For example, the typical vegetation spectral 834 curves in the images generated by the HyCoNet, SRGAN, and 835 WGAN reveal distinct biases in the range of red-edge to near-836 infrared with real HR data (see the images of the first row 837 in Fig. 9). In contrast, the vegetation spectral curves from the 838 LE-GAN super-resolution are more consistent with those from 839 real HR HSI. A detailed analysis of the spectral residual and 840 standard deviation between the generated HSI and real HR 841 HSI from the independent test dataset is shown in Fig. 10. 842 It can be found that the residual error between the LE-GAN 843 generated HSI and HR HSI is close to zero in the range of 450 844

853



Fig. 9. Detailed spectral analysis on local patches of false colour super resolution (\times 8) results generated by different models from the independent test dataset, GPF-3. The results in each column from the left to right are for real low resolution (LR) HSI patches, high resolution images generated from models (SRCNN, SRResNet, VDSR, SRGAN, WGAN, and the proposed LRE-GAN), and the corresponding High resolution (HR) HSI patches, respectively.

- to 780 nm and lower than 0.25 in the range of 780 to 950 nm, 845 and the deviation is lower than 0.02. All these results suggest 846 that the proposed model provides a better performance in HSI 847 super-resolution without losing the spectral details. The second 848 and third best spectral residuals are achieved by the BAGAN 849 and LTTR, respectively, and the spectral biases in the range 850 of 630 to 950 nm and the average deviations reach 0.029 and 851 0.041, respectively. 852

F. Experiment 3: Mode collapse evaluation

Generally, mode collapse mainly happens in the training 854 process when the super-resolution HSI produced by the gen-855 erator only partially covers the target domain. When the 856 generator learns a type of spectral-spatial pattern that is able 857 to fool the discriminator, it may keep generating this kind 858 of pattern so that the learning process is over-learned. The 859 distance and distribution of the generated super-resolution HSI 860 provide the most direct evidence for determining whether the 861 mode collapse occurs in the generator. In this section, we 862 evaluated the effect of the proposed LE-GAN on alleviating 863 mode collapse from three aspects: 1) a quantitative evaluation 864 on the diversity of the generated super-resolution HSI, based 865 on the distance-derived IS and FID metrics, 2) a smoothness 866 monitoring on the generator iterations during the network 867 training process, and 3) a visualisation of the distributions of 868 the real high-resolution HSI samples and the generated super-869 resolution samples. 870

Firstly, the quantitative evaluation for the diversity of the generated super-resolution HSI was conducted on the testing dataset and independent dataset mentioned in Section IV-C. In addition, in order to assess the potential affects of different upscaling factors and added-noise levels on the occurrence of mode collapse, all of the experiments were conducted on three



Fig. 10. A spectral residual (the black line) and deviation (the grey shadow) analysis between real HR HSI and the generated super-resolution HSI from different models on the model test dataset (KSC data) and the independent test dataset (GPF-3 data)

upscaling factors ($\times 2$, $\times 4$ and $\times 8$) with three Gaussian white 877 noise levels (∞ , 40db and 80db), and compared with five state-878 of-the-art competition models. The IS, FID, and Non-ref Score 879 were used as the evaluation metrics for assessing the diversity 880 of the super-resolution HSIs and determining the existence of a 881 mode collapse. A higher IS and lower FID and Non-ref Score 882 will show the better diversity of the generated super-resolution 883 HSI and the sign of the alleviation of mode collapse. 884

Table IV lists the IS, FID and Non-ref Score measurements 885 on the proposed LE-GAN and five selected competitors using 886 the testing datasets. The evaluation results on AVIRIS and 887 UHD-185 testing datasets demonstrate that the proposed LE-888 GAN model outperforms its competitors in terms of the IS, 889 FID, and Non-ref score measurements for all three upscaling 890 factors and added noise combinations (see the highlighted 891 values in Table IV). They also indicate that the proposed model 892 has greater performance on alleviating mode collapse issue 893 occurred in the generated spectral-spatial diversity. The dif-894 ferences of these measurements between the proposed model 895 and its competitors are particularly significant for the cases of 896 low added noise and low upscaling levels. For example, the 897 proposed LE-GAN achieves the highest IS (13.46 for AVIRIS 898 data and 14.69 for UHD-185 data), the lowest FID (13.7 for 899 AVIRIS data and 37.95 for UHD-185 data), and lowest Non-900 ref Score (11.17 for AVIRIS data and 19.31 for UHD-185 901 data) on the datasets with the $\times 2$ upscaling factor and non-902 added noise of ∞ SNR. 903

In addition, Table IV also reveals the IS, FID and Non-ref Score degradations with the increase of upscaling factor and TABLE IV A COMPARISON OF INCEPTION SCORES (IS), FRECHET INCEPTION DISTANCES (FID), AND NON-REFERENCED SPECTRAL SCORE (NON-REF SCORE) OF SUPER-RESOLUTION HSIS GENERATED FROM THE PROPOSED MODEL AND FIVE COMPETITION MODELS USING THE MODEL TEST

DATASETS.

			AVIRIS	5	UHD-185			
Upscaling	SNR		IS	FID	Non-ref Score	IS	FID	Non-ref Sco
		HyCoNet	11.63	57.37	21.42	9.64	80.15	36.65
		LTTR	11	45.55	25.31	8.61	78.67	37.23
	∞	BAGAN	12.26	48.81	18.12	12.01	70.34	31.25
		SRGAN	10.62	53.88	23.25	11.06	74.05	35.52
		WGAN	13.25	24.37	16.25	13.66	49.13	30.21
		LE-GAN	13.46	13.7	11.17	14.69	37.95	19.31
		HyCoNet	7.63	58.59	22.26	6.07	96.89	39.92
		LTTR	6.79	50.23	23.14	7.99	77.26	31.58
2	80	BAGAN	7.95	49.87	21.02	7.21	83.95	29.92
		SRGAN	6.77	54.59	25.24	6.09	92.06	30.32
		WGAN	11.37	24.45	15.35	8.12	60.83	24.35
		LE-GAN	11.56	15.86	15.21	12.16	40.34	20.82
		HyCoNet	4.35	63.01	26.68	4.05	104.3	41.23
		LTTR	5.02	52.42	27.24	4.9	80.12	40.56
	40	BAGAN	6.32	49.59	25.64	4.79	87.1	36.64
		SRGAN	5.04	60.55	28.66	4.61	94.97	34.25
		WGAN	9.17	26.71	19.92	5.67	76.48	29.25
		LE-GAN	10.16	18.94	17.81	10.13	49.53	25.56
		HyCoNet	9.79	62.14	27.25	8.87	87.65	35.58
		LTTR	9.82	49.46	22.45	7.1	86.22	34.45
	∞	BAGAN	11.03	53.51	26.24	10.69	77.25	27.69
		SRGAN	9.64	58.82	25.56	12.78	80.93	31.18
		WGAN	11.14	26.25	19.62	11.87	75.45	25.45
		LE-GAN	12.22	15.37	16.13	14.41	40.76	20.24
		HyCoNet	6.19	63.88	27.98	5.35	105.44	50.2
		LTTR	5.91	54.81	24.42	7.08	84.31	32.45
4	80	BAGAN	6.49	53.74	21.25	6.52	91.22	41.57
		SRGAN	5.4	58.94	26.58	5.42	101.1	51.5
		WGAN	9.89	26.71	20.62	7.35	77.32	29.8
		LE-GAN	11.28	17.84	17.99	12.77	43.86	24.85
		HyCoNet	3.83	68.34	35.9	2.91	113.66	55.97
		LTTR	4.32	57.13	27.41	4.49	87	38.52
	40	BAGAN	5.07	54.2	24.58	4.35	94.9	42.51
		SRGAN	4.39	66.08	33.85	3.72	103.29	52.04
		WGAN	7.55	28.64	20.14	5.22	83.75	32.24
		LE-GAN	10.43	19.91	18.38	10.31	50.53	27.25
		HyCoNet	8.44	64.71	34.12	8.04	91.34	48.82
		LTTR	8.97	52.06	28.78	6.53	89.81	45.58
	∞	BAGAN	9.33	55.47	25.54	9.31	81.3	38.88
		SRGAN	8.85	61.81	31.15	10.92	84.82	39.98
		WGAN	9.95	26.7	19.25	10.93	79.14	29.28
		LE-GAN	11.67	16.42	18.88	12.11	42.4	24.58
		HyCoNet	4.94	66.62	38.85	4.35	110.48	60.15
8		LITR	4.57	57.24	29.82	5.71	88.31	39.94
	80	BAGAN	5.96	55.75	26.57	5.3	95.53	46.25
		SRGAN	4.54	61.99	35.42	4.1	106.1	52.24
		WGAN	8.89	28	20.24	6.1	81.33	27.72
		LE-GAN	10.21	19.76	19.98	11.84	45.92	25.58
		HyCoNet	3.02	71.35	42.12	2.25	119.21	71.15
		LTTR	3.92	59.84	32.42	3.91	91.52	48.82
					21.14			
	40	BAGAN	4.32	56.21	51.14	5.12	98.9	53.35
	40	BAGAN SRGAN	4.32	56.21 69.28	39.92	2.69	98.9 108.21	53.35 59.94

added noise level for all of the models. The comparison of 906 these degradations can help to explore the model robustness 907 in preventing the mode collapse issue. Specifically, an average 908 of 33.6% IS drop, 32.03% FID increase, and 30.68% Non-909 ref score increase, are observed from the LE-GAN-based 910 super-resolution results when increasing the upscalling factor 911 from $\times 2$ to $\times 8$ and decreasing SNR from from ∞ to 40db. 912 Meanwhile, the results from the WGAN, the second best 913 model in terms of super-resolution fidelity (see Section IV-E), 914 show an average of 59.08% IS drop, 40.97% FID increase, 915 and 38.78% Non-ref Score increase. These findings suggest 916 that the proposed LE-GAN achieves the best performance in 917 preventing mode collapse under the higher upscaling factor 918 and noise interferences. 919

Table V provides the scores of the IS, FID, and Non-ref 920 Score from the proposed LE-GAN and its five competitors 921 using the independent datasets. Similar to the results shown 922 in Table IV, the results highlighted in Table V illustrate 923 that the LE-GAN model achieves the best and most robust 924 performance in terms of IS and FID for all the upscaling 925 factor and added noise combinations. In the case of a $\times 2$ 926 upscaling factor and non-added noise, the LE-GAN achieves 927 the best IS, FID, and Non-ref Score measurements (IS of 928 12.91, FID of 14.95, and Non-ref Score of 9.12 for AVIRIS 929 and IS of 15.27, FID of 39.22, and Non-ref Score of 19.99 930

TABLE V A Comparisons of Inception Scores (IS), Frechet Inception Distances (FID), and Non-referenced spectral score (Non-ref Score) from the proposed model and five competition models using the independent test datasets.

			5	UHD-185				
Upscaling	SNR		IS	FID	Non-ref Score	IS	FID	Non-ref Score
		HyCoNet	11.31	59.85	28.81	11.72	81.51	35.52
		LTTR	11.58	48.61	23.21	9.71	80.69	32.54
	∞	BAGAN	11.28	52.14	25.52	12.58	72.74	28.84
		SRGAN	10.91	57.35	28.15	13.25	76.32	26.62
		WGAN	12.22	26.26	14.21	14 45	70.71	22.82
		LEGAN	12.01	14.05	0.12	15 27	20.22	10.00
		HuCoNot	7 12	61.96	20.21	6.02	08.2	47.95
		ITYCONEL	6.07	62.77	30.21	0.95	70.2	47.85
2	0.0	LIIK	0.97	55.11	27.62	9.09	/9.18	38.43
2	80	BAGAN	8.32	52.03	27.41	8.48	85.1	39.99
		SKGAN	6.72	38.12	29.92	6.7	94.95	46.52
		WGAN	9.89	26.77	17.35	9.31	73.04	29.98
		LE-GAN	10.91	16.51	10.21	13.47	41.66	21.42
		HyCoNet	4.07	65.47	34.25	5.1	107.41	58.25
		LTTR	5.34	55.72	28.81	7.38	82.46	36.65
	40	BAGAN	5.34	52.91	26.65	5.56	88.16	39.94
		SRGAN	5.13	63.8	33.51	4.76	97.11	49.95
		WGAN	7.64	27.33	19.24	7.02	79.78	24.94
		LE-GAN	9.18	19.79	12.17	11.4	50.11	22.84
		HyCoNet	9.1	65.06	35.52	10.97	89.56	38.85
		LTTR	9.62	52.16	28.84	8.53	88.38	40.24
	∞	BAGAN	9.89	56.53	26.25	11.57	79.97	39.95
		SRGAN	9.46	62.35	31.54	13.65	83.64	33.35
		WGAN	10.41	27.73	18.82	14.56	77.62	27.74
		LE-GAN	10.76	14.8	11.15	14 18	42.2	21.22
		HyCoNet	5 49	67.05	38.88	6.27	107 44	59.94
		LTTR	6.07	58 44	30.45	7.45	86.43	30.88
4	80	BAGAN	6.68	56 30	20.04	7.26	03.15	43.25
-	00	SPGAN	5.00	62.05	25.52	6.04	102.04	50.05
		WCAN	0.74	28.27	10.04	0.04	70.41	20.00
		LECAN	0.74	19.25	19.94	0.0	19.41	29.99
		LE-GAN	9.94	10.35	12.25	11.04	45.02	22.51
		HyCoNet	3.85	/1.0/	45.25	4.20	116.89	69.94
	40	LITR	4.7	60.31	36.82	6.37	89.89	48.85
	40	BAGAN	5.61	57.67	29.94	3.85	96.17	57.74
		SRGAN	4.3	69.84	35.52	4.11	105.25	69.99
		WGAN	6.57	29.18	22.15	6.86	86.78	32.24
		LE-GAN	8.3	21.05	14.21	10.99	54.84	24.21
		HyCoNet	8.4	68.3	38.84	9.06	93.81	45.52
		LTTR	8.22	55.14	26.68	7.37	92.68	43.32
	∞	BAGAN	9.27	58.74	29.94	10.25	83.57	33.35
		SRGAN	8.31	65.12	36.65	11.83	87.25	39.94
		WGAN	9.31	28.2	18.74	12.29	81.61	31.15
		LE-GAN	9.8	15.84	12.29	12.82	43.43	23.01
		HyCoNet	4.53	70	39.98	5.47	113.15	63.35
		LTTR	4.61	60.32	28.84	6.73	90.34	40.41
8	80	BAGAN	5.58	58.78	27.74	5.78	97.65	51.15
5		SRGAN	4.51	65.9	35.58	5.13	108.75	57.74
		WGAN	8.81	29.62	21.82	6.85	83 75	32.25
		LE-GAN	9 17	18.05	13 35	10.16	47.84	24.61
		HuCoNat	2 27	75.03	12.55	2 22	122.07	27.01
		TTP	3.57	63 42	44.30	4.22	04.2	02.98
	40	PAGAN	2.09	50.71	33.38	4.23	94.2	49.63
	40	DAGAN	3.98	39.71	51.99	3.90	101.72	55.24
		SKGAN	3.03	75.26	40.65	5.54	110.76	08.95
		WGAN	6.7	31.34	22.25	5.76	90.11	39.98

for UHD-185 dataset), with the smallest IS drop (39.1%), FID increase (31.55%), and Non-ref Score increase (30.91%). These results are consistent with the mode collapse assessment reported in Table IV, suggesting that the LE-GAN derived super-resolution HSIs have the best spectral-spatial diversity with alleviated mode collapse.

Secondly, a smoothness monitoring on the generator itera-937 tion was used to determine if mode collapse occurred during 938 the training process. According to the results illustrated in 939 Table IV and Table V, the high noise-levels and the large 940 upscaling factors lead to more serious mode collapse. To 941 demonstrate the performance difference between the proposed 942 model and its competitors in alleviating model collapse, a 943 comparison was conducted under a high added noise level 944 (e.g. 40*db*) and a high upscaling factor (e.g. \times 8). Fig. 11 945 illustrates the IS and FID iterations of the generated HSIs 946 from the proposed LE-GAN and the two best competitors, 947 i.e. WGAN and BAGAN, during the training process. It is 948 obvious that the IS and FID curves from the proposed model 949 are smoother and more stable than those from WGAN or 950 BGAN, along with the increase of iteration number. Unlike 951 the curves from WGAN or BGAN, the curves of IS from the 952 proposed mode, LE-GAN, steadily increase and the curves of 953 FID steadily decrease for both AVIRIS and UHD-186 datasets. 954 This indicates that there is no significant mode collapse occurs 955



Fig. 11. The changes of IS and FID scores during the training of the proposed LE-GAN and its two best competitors (WGAN and BAGAN). The model training is conducted on (a-b) AVIRIS and (c-d) UHD-186 training dataset with a SNR level of 40db and an upscaling factor of $\times 8$.

during the training of LE-GAN. However, a big drop of IS 956 is observed during the training of BAGAN (e.g. after 3500 957 iterations, as shown in Fig. 11a) and during the training 958 of WGAN (e.g. after 2000 iterations, as shown in Fig. 11c). 959 Moreover, the curves of FID don't steadily decrease during the 960 training for WAGAN or BAGAN. These observations indicate 961 that the mode collapse occurs in the training of representative 962 GAN models (e.g. WGAN and BAGNAN), and the proposed 963 model is more effective in alleviating the mode collapse. 964

Finally, to further understand and assess the performance of 965 the generator model in dealing with the mode collapse issue, 966 the distributions of the real high-resolution HSI (I^{HR}) and the 967 generated super-resolution HSI (I^{SR}) were visualised in the 968 feature space, where the probability densities of the discrimi-969 nator eigenvalues of I^{HR} and I^{SR} , denoted as $D(I^{HR})$ and 970 $D(I^{SR})$, were used to represent the sample distributions. With 971 I^{HR} and I^{SR} as inputs separately fed into the discriminator 972 D described in Section III-A2, the outputs from the last 973 Maxpool layer, denoted as $D(I^{HR})$ and $D(I^{SR})$, represent the eigenvalues of the inputs I^{HR} and I^{SR} in the high-974 975 level discriminating space, and the probability densities of the 976 $D(I^{HR})$ and $D(I^{SR})$ represent the sample distributions of 977 I^{HR} and I^{SR} . The coverage of probability densities between 978 the $D(I^{HR})$ and $D(I^{SR})$ represent the mode similarity of the 979 I^{HR} and I^{SR} to indicate whether model collapse occurs in 980 the generator. 981

Fig. 12 illustrates that the probability density curves of 982 $D(I^{HR})$ and $D(I^{SR})$ obtained for three GAN models, LE-983 GAN and its two best competitors, WGAN and BAGAN, 984 through training the models on AVIRIS and UHD-185 datasets 985 with an SNR level of 40db and an upscaling factor of $\times 8$. In 986 comparison with the other two models, the probability density 987 curves of I^{SR} generated by LE-GAN are much closer to those 988 of the real I^{HR} for both AVIRIS (Fig. 12a) and UHD-185 989 datasets (Fig. 12d). However, the probability density curves of 990 the I^{SR} generated by WGAN (12b and e) and BAGAN (12c 991



Fig. 12. Statistic distributions of the high-resolution HSI (I^{hr}) and the generated super-resolution HSI (I^{SR}) in the discriminator network (D). We tested the proposed LE-GAN with the two best competitors (i.e. WGAN and BAGAN) on AVIRIS (a-c) and UHD-186 (d-f) test dataset with an SNR level of 40db and an upscaling factor of $\times 8$.

and f) have an obvious tendency shifting towards the right 992 and having a higher peak (i.e. a lower standard deviation). 993 This means the I^{SR} generated by WGAN or BAGAN can be 994 better discriminated from the real I^{HR} by D (i.e. low spectral-995 spatial fidelity), and the generated I^{SR} only covers the limited 996 spectral-spatial patterns of the real I^{HR} (i.e. existing the mode 997 collapse issue). These observations shows that the proposed 998 model outperforms the competitors in generating diversity of 999 super-resolution samples and alleviating mode collapse. 1000

V. DISCUSSION

1001

The challenge of GANs in improving the spectral and spatial 1002 fidelity of HSI super-resolution and addressing the issue of 1003 mode collapse is on how to make the generator learn the real 1004 spectral-spatial patterns, and meanwhile, prevent the generator 1005 from over-learning limited patterns. Since there is no such 1006 kinds of constraints in the JS distance based loss functions, the 1007 original GAN is hard to generate the high fidelity HSI super-1008 resolution and easy to suffer mode collapse. In this study, we 1009 proposed a novel GAN model, named as LE-GAN, through 1010 improving the GAN baseline and introducing a new SSRP 1011 loss function. The new SSRP loss was used to guide the 1012 optimisation and alleviate the spectral-spatial mode collapse 1013 issue occurred in the HSI super-resolution process. The model 1014 validation and evaluation were conducted using the datasets 1015 from two hyperspectral sensors (i.e. AVIRIS and UHD-185) 1016 with various upscaling factors ($\times 2$, $\times 4$, and $\times 8$) and added-1017 noises (∞ db, 40db, and 80db). The evaluation results showed 1018 that the proposed LE-GAN can achieve high-fidelity HSI 1019 super-resolution for relatively high upscaling factors and have 1020 a better robustness against noise and better generalizability to 1021 various sensors. 1022

1023 A. The ablation analysis of the improved modules

In the proposed model, a total of five different modifications have been made to improve the GAN baseline including: 1) using 3D-convolutional filters in G, 2) adding an Upscale-Block in G, 3) removing the sigmoid in D, 4) adding a novel La network, and 5) using a new loss function to optimise the model. 1028

To evaluate the effects of these improvements on the perfor-1030 mance of the proposed LE-GAN, we have conducted an abla-1031 tion analysis in which we gradually substituted the traditional 1032 GAN components with the proposed modules and compares 1033 their effects based on six evaluation metrics, PSNR, SIM, PI, 1034 SAM, SRE, and computing time (CT). Each improvement is 1035 an incremental modification to the original GAN model, thus 1036 forming five different models: Model 1 to Model 5. The details 1037 of the five models and their influences on the six evaluation 1038 metrics for the testing datasets (AVIRIS and UHD-185) with 1039 $\times 8$ scale factor are presented in Fig.13. The super-resolution 1040 results of three example patches are also displayed for the 1041 visual comparison. 1042

1) Model 1: using 3D-convolutional filters in G: In order to process continuous spectral channels and capture spectralspatial joint features learning in the ResBlock in G, 3Dconvolutional filters are used. Theoretically, this modification is able to extract both the spectral correlation characteristics and spatial texture information.

2) Model 2: Adding an UpscaleBlock in G: In a super-1049 resolution network, the most important strategy to improve 1050 the performance is to increase the information (e.g. the di-105 mensionality of feature maps) of an LR HSI to match with 1052 that of the corresponding HR HSI. However, the traditional 1053 approaches increase the feature dimensionality in the entire 1054 intermediate layers gradually, which increases the computation 1055 complexity and computational cost. In contrast, we proposed 1056 an UpscaleBlock to super-resolve the detailed spectral-spatial 1057 information only at the end of the generator network (see 1058 Fig. 3). This adjustment directly eliminates the need of the 1059 computational and memory resources for super-resolution op-1060 erations. Thus, a smaller filter size can be effectively used in 1061 our generator network for the extraction of super-resolution 1062 features. The results of Model 2 (the third column in Fig. 1063 13) reveals a performance improvement after adding the 1064 UpscaleBlock. Compared to Model 1, the computation time 1065 has a 35.1% reduction on average without losing the super-1066 resolution quality, Model 2 even has a better super-resolution 1067 quality in terms of PSNR, SSIM, PI, SAM and SRE. 1068

3) Model 3: Removing the sigmoid function from the dis-1069 criminator: In the traditional GAN framework, the sigmoid-1070 activated features often skew the original feature distribution 107 and result in lower reconstructed spectral values. Therefore, in 1072 this study, we removed the sigmoid activation in D network 1073 for two reasons. Firstly, using the feature before activation 1074 can benefit accurate reconstruction of the spectral and spatial 1075 features of input. Secondly, the proposed latent space distance 1076 requires real feature distribution of the input HSI in the low-1077 dimensional manifold in order to measure the divergence 1078 between the generated super-resolution HSI and real HR HIS. 1079 This modification, as shown in Model 3 in the fourth column 1080 of Fig. 13, contributes to an approximately 7.2% reduction 1081 in SAM and 13.9% reduction in SRE. These findings suggest 1082 that removing the sigmoid activation can help keep the spectral 1083 consistency between the LR and HR HSIs. 1084

4) Model 4: Adding a newly developed L_E network: The 1085 L_E network is developed to produce a latent regularisation 1086 term, which holds up the manifold space of the generated 1087 super-resolution HSI so that the dimensionality of the gen-1088 erated HSI is consistent with that of real HR HIS. In addition, 1089 the L_E network makes the divergence of the generated HSI 1090 and real HSI satisfy the Lipschitz condition for optimisation. 1091 The generated super-resolution HSI patches from Model 4 (see 1092 the fifth column of Fig. 13) indicates that, after adding the L_E 1093 network into the original GAN framework, both SAM and 1094 SRE have a significant reduction, with a drop of 6.7% and 1095 15.3%, respectively. Besides, there is a slight improvement on 1096 the PNSR, SSIM, and PI (the PNSR and SSIM respectively 1097 increase 2.8% and 1.4%, the PI declines 4.3%). These results 1098 indicate that the regularisation term produced by the L_E 1099 network has a great contribution in reconstructing the spectral-1100 spatial details consistent with real HR HIS. However, the L_E 1101 need to occupy a certain amount of computational and memory 1102 resources, subsequently the computation time increases 15.1%. 1103

5) Model 5: Using the new loss function to optimise the 1104 *model:* The most important contribution of our work is to 1105 develop a SSRP loss function with a latent regularisation to 1106 optimise the whole model. Model 5 (see the last column of 1107 Fig. 13), the final version of the LE-GAN, improves all of 1108 the evaluation metrics. The increases of PNSR and SSIM are 1109 5.2% and 12.4%, respectively, while the decreases of SAM 1110 and SRE are 13.1% and 7.9%, respectively. But, it leads to a 1111 11.4% increase of computation time. These findings suggest 1112 that the proposed SSEP loss function with the latent space 1113 regularisation can boost the performance on measuring the 1114 divergence of generated HSI and real HSI in both spectral 1115 and spatial dimensionality. 1116

1117 B. The Evaluation of the loss function

The proposed loss function introducing latent regularisation 1118 into the Wasserstein loss function optimises the GAN in the 1119 1120 latent manifold space and addresses the problems of mode collapse. In order to verify the effectiveness of the proposed 1121 loss function, we trained the proposed LE-GAN model with 1122 three kinds of losses: 1) the traditional JS divergence-based 1123 loss, 2) the Wasserstein distance-based loss, and 3) the pro-1124 posed improved Wasserstein loss with latent regularisation, 1125 and plotted their loss curves on both the training and validation 1126 sets in Fig. 14. 1127

It is obvious that the training process of the model with 1128 JS divergence-based loss, as shown in Fig. 14a, is un-1129 а stable and volatile. The reason behind lies in the fact that 1130 the JS divergence always leads to the supports of P_r and 1131 P_q disjointing in the low-dimensional manifolds during the 1132 process of maximising the discriminative capability of D, 1133 which causes the gradient fluctuation. On the contrary, the 1134 Wasserstein distance based loss functions, as shown in Fig. 1135 14b and c, can improve the stability of learning and lead 1136 the loss converges to the minimum. This findings is con-1137 sistent with Arjovsky et al. [50]'s and Ishaan et al. [51]'s 1138 studies. In addition, it is noteworthy that the loss curve of 1139 the proposed model is more stable and smoother than that of 1140



PSNR/SSIM/PI/ 34.21/0.91/4.9 34.81/0.92/4.81 34.91/0.92/4.77 35.14/0.92/4.39 37.52/1.04/4.11 SAM/SRE/CT /8.78/14.62/89.3s /8.71/13.6/56.17s /8.61/13.02/53.9s /7.95/10.36/60.4s /6.99/9.72/63.5s

Fig. 13. The influence of five incremental modifications to the performance of the proposed model in terms of PSNR, SSIM, PI, SRE and computing time (CT). The results in each column (except first column) correspond to one new model with one incremental change to its previous model. The superresolution results of different models with $\times 8$ scale factor on the testing datasets are illustrated for visual comparison.

the traditional Wasserstein distance-based losses. The theory 1141 behind is that introducing the latent regularisation terms into 1142 the training process provides a non-singular support to the 1143 generated sample sets at the corresponding low-dimensional 1144 manifolds. It is expected that the Wasserstein distance (i.e. 1145 $W(P_r, P_q)$) performs better under the condition of the con-1146 tinuity and differentiability of the divergence of P_r and P_q). 1147 With the latent regularisation, the max-min game of LE-GAN 1148 will yield a probability distribution $P_q(G(I^{lr}))$ in a low-1149 dimensional manifold that has a joint support with $P_r(I^{hr})$, 1150 and the process of minimizing the $W(P_r, P_q)$ will facilitate 1151 the gradient descent of the trainable parameters in G because 1152 the valid gradient can be captured from the optimised D in the 1153 low dimensional manifold. Therefore, the latent regularisation 1154 derived Wasserstein loss is regarded as a more sensible loss 1155 function for HSI super-resolution than the JS divergence loss 1156 and the traditional Wasserstein loss. 1157

The subplots above the learning curve shown in Fig. 14 are 1158 the images generated in the optimisation process when three 1159 different losses are used. It is obvious that the super-resolved 1160 HSI subplots optimised by the JS divergence-based loss (see 1161 Fig. 14a) do not produce the equivalent quality of spatial 1162 texture reconstruction as those from the proposed model (see 1163 Fig. 14b). The proposed latent regularisation term makes the 1164 dimensionality of the generated HSI manifold more consistent 1165 with that of the HR HSI in the optimisation process. 1166



Fig. 14. A comparison of the loss curves during the training of Model 4 using (a) the traditional JS divergence-based loss (b) the Wasserstein distance-based loss, and (c) the proposed improved Wasserstein loss with latent regularisation.

1167 C. Model robustness analysis

Our experimental results have illustrated the great quality 1168 scores and robustness for the proposed LE-GAN model. To 1169 further analysis the robustness of our proposed model on 1170 resisting the complicated down-sampling blurring in practical 1171 applications, we discuss the potential of our model gener-1172 ality from two aspects: 1) the theoritical analysis and 2) 1173 the additonal testing and evaluation on on real AVIRIS data 1174 products. In the natural conditions, the downsampling bluring 1175 is generally caused by two factors, atmospheric effects (e.g. the 1176 Gaussian noise and mixed Gaussian noises), and instrumental 1177 (sensor) noises (e.g. salt & pepper noise and speckle noise 1178 etc). Because the instrumental noises, as part of the internal 1179 system errors, are decided by the hardware parameters of a 1180 sensor (e.g. the band setting and SNR), here we only discuss 1181 the atmospheric effects (i.e. Gaussian noises) under the real 1182 natural conditions. 1183

Mathematically, the observed reflectance $(R_{observation})$ can 1184 be formulated as the sum of the real reflectance (R_{real}) 1185 and noises (N), thus $R_{observation} = R_{real} + N$. Based 1186 on our model setting, the spectral-spatial features (SSF) of 1187 the $R_{observation}$ will be extracted, and then, the singular 1188 value decomposition of spectral-spatial features (SSF) will be 1189 conducted in the latent encoder as shown in equation 4. After 1190 that, the spectral-spatial distribution (SSD) of SSF used for 1191 training the latent regularisation term can be expressed as the 1192 sum of the SSD of the real reflectance features (SSD_{real}) 1193 and the noise features (SSD_N) . Therefore, it is obvious that 1194

1210

1245

the influence of the noise in our model depends on the rank 1195 of SSD_N , thus the entropy of the noise, rather than the type of noise. This shows that our model has great potential 1197 in tolerating the complicated noise corruptions of HSI. This conclusion is also proven by our experimental results in Section IV-E.

In addition, to further test the generality of our model 1201 on more real HSI, we applied our model on the 20 real 1202 AVIRIS HSI dataset downloaded from the websit (https : 1203 //aviris.jpl.nasa.gov/dataportal/). The results were eval-1204 uated in terms of PSNR, SSIM, PI, SAM, SRE, IS, and FID 1205 provided in Table S1 of the supplementary. These results are 1206 consistent with our experimental results reported in Section 1207 IV-E. It concludes that the proposed model has great robust-1208 ness and generality for the practical applications. 1209

D. Limitations and future works

Benefiting from the self-adaptive latent encoder, the pro-1211 posed LE-GAN architecture performs better on alleviating 1212 the spectral-spatial distortion caused by the mode collapse 1213 issue than traditional GAN networks, and the robustness and 1214 generalizability of the proposed LE-GAN for HSI super-1215 resolution with large upscaling factor and higher noise levels 1216 are better than existing models. The experimental evaluation 1217 has illustrated a good spectral-spatial fidelity and diversity 1218 of the HSI super-resolution generated by the proposed LE-1219 GAN. Our model will be more adaptive in real applications 1220 with different disturbances of HSI super-resolution since the 1221 proposed method introduces the STSSRW mechanism into the 1222 generator network to enhance the hierarchical spectral-spatial 1223 information and ignore the non-hierarchical noises during 1224 the upscaling processing. However, there are two limitations 1225 should be noticed. First, because our current input data for 1226 model training only includes two types of sensors (i.e. UHD-1227 185 and AVIRIS), the direct use of this pre-trained model may 1228 lead to limited performance on the unseen images from other 1229 sensors with different spectral band settings. Second, in the 1230 proposed LE-GAN architecture, an extra latent encoder net-1231 work will inevitably introduce a lot of parameters and increase 1232 the computational complexity. Based on these two limitations, 1233 our future work will focus on improving the performance 1234 of our model by tackling these two challenges: 1) we will 1235 develop a Reinforcement Learning (RL) strategy to utilize 1236 the knowledge gained in our pre-trained model in unseen 1237 novel tasks, and further test and evaluate the generalization 1238 capacities of the proposed adversarial training methods in 1239 more complicated and unexpected conditions, 2) Considering 1240 the training processes of the generator, discriminator and latent 1241 encoder are relatively independent, we will develop a parallel 1242 processing strategy to improve the computational efficiency of 1243 model training and testing. 1244

VI. CONCLUSION

To address the challenge of spectral-spatial distortions 1246 caused by mode collapse during the optimisation process, 1247 this work has developed a latent encoder coupled GAN for 1248 spectral-spatial realistic HSI super-resolution. In the proposed 1249

GAN architecture, the generator is designed based on an 1250 STSSRW mechanism with a consideration of spectral-spatial 1251 hierarchical structures during the upscaling process. In addi-1252 tion, a latent regularised encoder is embedded in the GAN 1253 framework to map the generated spectral-spatial features into 1254 a latent manifold space and make the generator a better 1255 estimation of the local spectral-spatial invariances in the 1256 latent space. For the model optimisation, a SSRP loss has 1257 been introduced to avoid the spectral-spatial distortion in the 1258 super-resolution HSI. By using the SSRP loss, both spectral-1259 spatial perceptual differences and adversarial loss in latent 1260 space are measured during the optimization process. More 1261 importantly, a latent regularisation component is coupled with 1262 the optimisation process to maintain the continuity and no-1263 singularity of the generated spectral-spatial feature distribution 1264 in the latent space and increase the diversity of the super-1265 resolution features. We have conducted different experimental 1266 evaluation in terms of mode collapse and performance. The 1267 proposed approach has been tested and validated on AVIRIS 1268 and UHD-185 HSI datasets and compared with five state-of-1269 the-art super resolution methods. The results show that the 1270 proposed model outperforms the existing methods and is more 1271 robust to noise and less sensitive to the upscaling factor. The 1272 proposed model is capable of not only generating high quality 1273 super-resolution HSIs (both the spatial texture and spectral 1274 consistency) but also alleviating mode collapse issue. 1275

ACKNOWLEDGMENT

1276

1286

This research was supported BBSRC (BB/R019983/1), BB-1277 SRC (BB/S020969/1). The work is also supported by Newton 1278 Fund Institutional Links grant, ID 332438911, under the 1279 Newton-Ungku Omar Fund partnership (the grant is funded 1280 by the UK Department of Business, Energy, and Industrial 1281 Strategy (BEIS)) and the Open Research Fund of Key Lab-1282 oratory of Digital Earth Science, Chinese Academy of Sci-1283 ences(No.2019LDE003). For further information, please visit 1284 www.newtonfund.ac.uk. 1285

REFERENCES

- [1] L. Deng, J. Sun, Y. Chen, H. Lu, F. Duan, L. Zhu, and T. Fan, "M2h-net: A reconstruction method for hyperspectral remotely sensed imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 323–348, 2021.
- [2] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler,
 "Super-resolution of sentinel-2 images: Learning a globally applicable
 deep neural network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 146, pp. 305–319, 2018.
- [3] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 639– 643, 2017.
- [4] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29– 56, 2017.
- [5] P. Ghamisi, B. Rasti, N. Yokoya, Q. Wang, B. Hofle, L. Bruzzone, F. Bovolo, M. Chi, K. Anders, R. Gloaguen *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 1, pp. 6–39, 2019.
- [6] J. Qu, Y. Li, and W. Dong, "Hyperspectral pansharpening with guided filter," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2152–2156, 2017.

- [7] G. Khademi and H. Ghassemian, "A multi-objective componentsubstitution-based pansharpening," in 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA). IEEE, 2017, pp. 248–252.
- [8] S. E. Brezini, M. S. Karoui, F. Z. Benhalouche, Y. Deville, and
 A. Ouamri, "Hypersharpening by an nmf-unmixing-based method addressing spectral variability," *IEEE Geoscience and Remote Sensing Letters*, 2021.
- C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral superresolution with spectral unmixing constraints," *Remote Sensing*, vol. 9, 1320 no. 11, p. 1196, 2017.
- [10] J. A. Benediktsson, G. Cavallaro, N. Falco, I. Hedhli, V. A. Krylov,
 G. Moser, S. B. Serpico, and J. Zerubia, "Remote sensing data fusion:
 Markov models and mathematical morphology for multisensor, multires olution, and multiscale image classification," in *Mathematical Models* for Remote Sensing Image Processing. Springer, 2018, pp. 277–323.
- T. Wang, F. Fang, F. Li, and G. Zhang, "High-quality bayesian pansharpening," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 227–239, 2018.
- J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, 1330
 N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 2, pp. 6–36, 2013.
- W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans-actions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.
- [14] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.
- [15] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, 1340
 A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1343 2017, Conference Proceedings, pp. 4681–4690.
- [16] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced gan for remote sensing image superresolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5799–5812, 2019.
- B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, Conference Proceedings, pp. 136–144.
- [18] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, Conference Proceedings, pp. 0–0.
- [19] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, Conference Proceedings, pp. 2472–2481.
- [20] Y. Xiong, S. Guo, J. Chen, X. Deng, L. Sun, X. Zheng, and W. Xu, "Improved srgan for remote sensing image superresolution across locations and sensors," *Remote Sensing*, vol. 12, no. 8, p. 1263, 2020.
- W. Li, L. Fan, Z. Wang, C. Ma, and X. Cui, "Tackling mode collapse in multi-generator gans with orthogonal vectors," *Pattern Recognition*, vol. 110, p. 107646, 2020.
- Y. Hou and J. Zhang, "Unsupervised remote sensing image superresolution method based on adaptive domain distance measurement network," in 2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE). IEEE, 2020, pp. 256–259.
- W. Ma, Z. Pan, F. Yuan, and B. Lei, "Super-resolution of remote sensing images via a dense residual generative adversarial network," *Remote Sensing*, vol. 11, no. 21, p. 2578, 2019.
- [24] K. Jiang, Z. Wang, P. Yi, J. Jiang, G. Wang, Z. Han, and T. Lu,
 "Gan-based multi-level mapping network for satellite imagery superresolution," in 2019 IEEE International Conference on Multimedia and Expo (ICME). IEEE, Conference Proceedings, pp. 526–531.
- [25] U. B. Gewali, S. T. Monteiro, and E. Saber, "Spectral super-resolution using with optimized bands," *Remote Sensing*, vol. 11, no. 14, p. 1648, 2019. 1379
- [26] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5046–5063, 2018.
- [27] R. Jiang, X. Li, A. Gao, L. Li, H. Meng, S. Yue, and L. Zhang, "Learning spectral and spatial features based on generative adversarial network for hyperspectral image super-resolution," in *IGARSS 2019*-1386

1387

1388

2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, Conference Proceedings, pp. 3161-3164.

- J. Li, R. Cui, Y. Li, B. Li, Q. Du, and C. Ge, "Multitemporal 1389 [28] hyperspectral image super-resolution through 3d generative adversarial 1390 network," in 2019 10th International Workshop on the Analysis of 1391 1392 Multitemporal Remote Sensing Images (MultiTemp). IEEE. 2019. Conference Proceedings, pp. 1-4. 1393
- Q. Huang, W. Li, T. Hu, and R. Tao, "Hyperspectral image super-1394 [29] resolution using generative adversarial network and residual learning," in 1395 ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech 1396 1397 and Signal Processing (ICASSP). IEEE, Conference Proceedings, pp. 3012-3016. 1398
- [30] G. Freedman and R. Fattal, "Image and video upscaling from local self-1399 examples," ACM Transactions on Graphics (TOG), vol. 30, no. 2, pp. 1400 1401 1-11.2011.
- [31] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, 1402 1403 D. Rueckert, and Z. Wang, "Real-time single image and video superresolution using an efficient sub-pixel convolutional neural network," in 1404 Proceedings of the IEEE conference on computer vision and pattern 1405 1406 recognition, 2016, Conference Proceedings, pp. 1874-1883.
- [32] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian 1407 pyramid networks for fast and accurate super-resolution," in Proceedings 1408 of the IEEE conference on computer vision and pattern recognition, 1409 1410 Conference Proceedings, pp. 624–632.
- [33] A. Rangnekar, N. Mokashi, E. Ientilucci, C. Kanan, and M. Hoffman, 1411 1412 "Aerial spectral super-resolution using conditional adversarial networks." arXiv preprint arXiv:1712.08690, 2017. 1413
- [34] P. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "Cnn-1414 based super-resolution of hyperspectral images," IEEE Transactions on 1415 Geoscience and Remote Sensing, 2020. 1416
- G.-J. Qi, "Loss-sensitive generative adversarial networks on lipschitz 1417 [35] 1418 densities," International Journal of Computer Vision, pp. 1-23, 2019.
- P. Samangouei, M. Kabkab, and R. Chellappa, "Defense-gan: Protecting 1419 [36] classifiers against adversarial attacks using generative models," arXiv 1420 preprint arXiv:1805.06605, 2018. 1421
- K. Liu, Q. Li, and G. Qiu, "Posegan: A pose-to-image translation 1422 [37] framework for camera localization," ISPRS Journal of Photogrammetry 1423 and Remote Sensing, vol. 166, pp. 308-315, 2020. 1424
- [38] H. Zhang, J. Ma, C. Chen, and X. Tian, "Ndvi-net: A fusion network 1425 for generating high-resolution normalized difference vegetation index 1426 in remote sensing," ISPRS Journal of Photogrammetry and Remote 1427 Sensing, vol. 168, pp. 182-196, 2020. 1428
- [39] W. Jiang, Z. Liu, K.-H. Lee, S. Chen, Y.-L. Ng, Q. Dou, H.-C. Chang, 1429 1430 and K.-W. Kwok, "Respiratory motion correction in abdominal mri using a densely connected u-net with gan-guided training," arXiv preprint 1431 arXiv:1906.09745, 2019. 1432
- S. K. Yarlagadda, D. Güera, P. Bestagini, F. Maggie Zhu, S. Tubaro, [40] 1433 and E. J. Delp, "Satellite image forgery detection and localization using 1434 gan and one-class classifier," Electronic Imaging, vol. 2018, no. 7, pp. 1435 214-1, 2018. 1436
- A. Bhattacharjee, S. Banerjee, and S. Das, "Posix-gan: Generating multi-1437 [41] 1438 ple poses using gan for pose-invariant face recognition," in Proceedings 1439 of the European Conference on Computer Vision (ECCV), 2018, pp. 0–0.
- X. Chen, C. Xu, X. Yang, and D. Tao, "Attention-gan for object trans-[42] 1440 figuration in wild images," in Proceedings of the European Conference 1441 on Computer Vision (ECCV), 2018, pp. 164-180. 1442
- M. Zhang and Q. Ling, "Supervised pixel-wise gan for face super-[43] 1443 resolution," IEEE Transactions on Multimedia, 2020. 1444
- K. Lei, M. Qin, B. Bai, G. Zhang, and M. Yang, "Gcn-gan: A non-linear 1445 [44] 1446 temporal link prediction model for weighted dynamic networks," in IEEE INFOCOM 2019-IEEE Conference on Computer Communications. 1447 IEEE, 2019, pp. 388-396. 1448
- [45] M. Cha, Y. Gwon, and H. Kung, "Adversarial nets with perceptual losses 1449 for text-to-image synthesis," in 2017 IEEE 27th International Workshop 1450 1451 on Machine Learning for Signal Processing (MLSP). IEEE, 2017, pp. 1-6.1452
- 1453 [46] X. Luo, R. Chen, Y. Xie, Y. Qu, and C. Li, "Bi-gans-st for perceptual image super-resolution," in Proceedings of the European Conference on 1454 Computer Vision (ECCV), 2018, pp. 4-6. 1455
- [47] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in Pro-1456 ceedings of the IEEE Conference on Computer Vision and Pattern 1457 1458 Recognition, 2018, pp. 6228-6237.
- [48] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel, and 1459 J.-P. Thiran, "Srobb: Targeted perceptual loss for single image super-1460 resolution," in Proceedings of the IEEE International Conference on 1461 Computer Vision, 2019, pp. 2710-2719. 1462

- [49] P. Bojanowski, A. Joulin, D. Lopez-Paz, and A. Szlam, "Optimizing the 1463 latent space of generative networks," arXiv preprint arXiv:1707.05776, 1464 2017 1465
- [50] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative ad-1466 versarial networks," ser. Proceedings of Machine Learning Research, 1467 D. Precup and Y. W. Teh, Eds., vol. 70. International Convention 1468 Centre, Sydney, Australia: PMLR, 06-11 Aug 2017, pp. 214-223 1469
- [51] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, 1470 "Improved training of wasserstein gans," in Advances in neural infor-1471 mation processing systems, Conference Proceedings, pp. 5767-5777. 1472
- [52] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering 1473 manifold structure for hyperspectral imagery super-resolution," IEEE 1474 Transactions on Image Processing, vol. 27, no. 12, pp. 5969-5982, 2018. 1475
- [53] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral 1476 image super-resolution," IEEE Transactions on Neural Networks and 1477 Learning Systems, vol. 32, no. 6, pp. 2388-2400, 2020. 1478
- S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional [54] 1479 neural network for dynamic scene deblurring," in Proceedings of the 1480 IEEE conference on computer vision and pattern recognition, 2017, pp. 1481 3883-3891 1482
- J. Cao and J. Wang, "Absolute exponential stability of recurrent neural [55] 1483 networks with lipschitz-continuous activation functions and time delays," 1484 Neural networks, vol. 17, no. 3, pp. 379-390, 2004. 1485
- [56] K. Zheng, L. Gao, W. Liao, D. Hong, B. Zhang, X. Cui, and J. Chanus-1486 sot, "Coupled convolutional neural network with adaptive response 1487 function learning for unsupervised hyperspectral super resolution," IEEE 1488 Transactions on Geoscience and Remote Sensing, vol. 59, no. 3, pp. 1489 2487-2502, 2020. 1490
- [57] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank represen-1491 tation for hyperspectral image super-resolution," IEEE transactions on 1492 neural networks and learning systems, vol. 30, no. 9, pp. 2672-2683, 1493 2019. 1494
- [58] J. Li, R. Cui, B. Li, R. Song, Y. Li, Y. Dai, and O. Du, "Hyperspectral 1495 image super-resolution by band attention through adversarial learning," 1496 IEEE Transactions on Geoscience and Remote Sensing, 2020. 1497
- [59] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral 1498 image spatial super-resolution via 3d full convolutional neural network," Remote Sensing, vol. 9, no. 11, p. 1139, 2017. 1500
- [60] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in 1501 2010 20th International Conference on Pattern Recognition. IEEE. 1502 Conference Proceedings, pp. 2366-2369.
- [61] S. S. Channappayya, A. C. Bovik, and R. W. Heath Jr, "Rate bounds 1504 on ssim index of quantized images," IEEE Transactions on Image 1505 Processing, vol. 17, no. 9, pp. 1624–1639, 2008.
- [62] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, 1507 "The 2018 pirm challenge on perceptual image super-resolution," in 1508 Proceedings of the European Conference on Computer Vision (ECCV), 1509 Conference Proceedings, pp. 0-0. 1510
- [63] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference 1511 quality metric for single-image super-resolution," Computer Vision and 1512 Image Understanding, vol. 158, pp. 1-16, 2017. 1513
- A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely [64] 1514 blind" image quality analyzer," IEEE Signal Processing Letters, vol. 20, 1515 no. 3, pp. 209-212, 2012. 1516
- [65] K. Liu, W. Tang, F. Zhou, and G. Qiu, "Spectral regularization for 1517 combating mode collapse in gans," in Proceedings of the IEEE/CVF 1518 International Conference on Computer Vision, 2019, pp. 6382–6390. 1519
- [66] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "Eeg-gan: Generative 1520 adversarial networks for electroencephalograhic (eeg) brain signals," 1521 arXiv preprint arXiv:1806.01875, 2018. 1522
- J. Yang, Y. Zhao, C. Yi, and J. C.-W. Chan, "No-reference hyperspectral [67] 1523 image quality assessment via quality-sensitive features learning," Remote 1524 Sensing, vol. 9, no. 4, p. 305, 2017. 1525
- G. Anbarjafari and H. Demirel, "Image super resolution based on [68] 1526 interpolation of wavelet domain high frequency subbands and the spatial 1527 domain input image," ETRI journal, vol. 32, no. 3, pp. 390-394, 2010. 1528
- [69] M. Srivastava, S. Pallavi, S. Chandra, and G. Geetha, "Comparison 1529 of optimizers implemented in generative adversarial network (gan)," 1530 1531 International Journal of Pure and Applied Mathematics, vol. 119, no. 12, pp. 16831-16836, 2018. 1532

1499

1503