Communication

Synthesis of Synthetic Hyperspectral Images with Controllable Spectral Variability Using a Generative Adversarial Network

Burkni Palsson, Magnus O. Ulfarsson and Johannes R. Sveinsson *

Faculty of Electrical and Computer Engineering, University of Iceland, 105 Reykjavik, Iceland; burkni@hi.is (B.P.); mou@hi.is (M.O.U.)
* Correspondence: sveinsso@hi.is

Abstract: In hyperspectral unmixing (HU), spectral variability in hyperspectral images (HSIs) is a major challenge which has received a lot of attention over the last few years. Here, we propose a method utilizing a generative adversarial network (GAN) for creating synthetic HSIs having a controllable degree of realistic spectral variability from existing HSIs with established ground truth abundance maps. Such synthetic images can be a valuable tool when developing HU methods that can deal with spectral variability. We use a variational autoencoder (VAE) to investigate how the variability in the synthesized images differs from the original images and perform blind unmixing experiments on the generated images to illustrate the effect of increasing the variability.

Keywords: hyperspectral unmixing; synthetic hyperspectral images; generative adversarial network; variational autoencoder; deep learning; neural network

1. Introduction

Endmember variability poses a significant challenge for linear unmixing methods that use scale-invariant objective functions. The linear mixing model (LMM) models the spectrum of a pixel $x_p$ for an HSI $Y \in \mathbb{R}^{P \times B}$ having $P$ pixels and $B$ bands as a convex combination of the endmembers:

$$x_p = As_p + \epsilon_p, \text{ subject to } 1^T s_p = 1 \text{ and } s_p \geq 0,$$

where the abundance fractions $s_p$ are the area proportions of the endmembers in the pixel and $\epsilon_p$ is the noise. The matrix $A$ is the endmember matrix, having the endmembers as its columns. When looking at Equation (1), we see that the LMM cannot model the spectral variability, which is a significant limitation. In remote sensing, the spectral variability is mainly caused by atmospheric effects, variation in illumination and topography, and intrinsic variations in the spectral signatures of the material itself [1]. Considerable effort has been put into developing HU techniques that handle spectral variability better.

Dealing with spectral variability has mainly been accomplished in two ways. One is using bundles of endmembers to represent each endmember class, with the bundles acquired from the data themselves or through field campaigns [2]. The other way is modeling the endmember variability using physical or statistical modeling. This requires extending the LMM in such a way that the endmembers can be different for each pixel. Some examples of such extended models are the augmented LMM [3], the extended LMM [4], and the perturbed LMM [5].

Numerous datasets exhibit two deficiencies when it comes to examining the impacts of spectral variability on methods. First, real datasets do not possess an easily quantifiable degree of spectral variability. Secondly, generated datasets typically have either zero spectral variability, or it is frequently produced based on a straightforward model, such as an augmented or extended linear mixing model [3,4], which makes the spectral variability somewhat rudimentary. It should be noted that there are advanced and complex methods,
such as those in [6], that can be used to simulate HSIs with physical simulations using advanced radiometric and geometric models.

Recently, deep generative models have been employed in the generation of HSIs. By learning the distribution of actual images, they can generate new samples from the same distribution. One example is in [7], where a generative adversarial network (GAN) [8] was utilized to create realistic synthetic HSIs for data augmentation in hyperspectral classification tasks. Another significant type of deep generative model is the variational autoencoder (VAE) [9], which employs a different method to learn the data distribution and generate new samples.

VAEs are especially adept at modeling complex, high-dimensional data and continuous latent spaces, making them extremely useful for tasks like generating diverse sets of similar images. VAEs have been used for a wide range of applications, including image processing, data augmentation [10], image generation [11], game design [12], and hyperspectral unmixing. An instance of using a VAE for generating HSIs was demonstrated in [13], where a VAE was used to create synthetic HSIs with adjustable spectral variability. Diffusion models [14] present yet another distinct approach to data generation. The concept behind diffusion models is to portray the data generation process as a diffusion procedure that incrementally introduces noise to the original data over time, converting it into a simple distribution. To produce new data, the process is reversed; initiating from noise, the model applies a series of transformations that progressively eliminate the noise, resulting in a sample from the data distribution. These transformations are guided by a neural network, which is trained to predict how to denoise a data sample at each stage of the diffusion process.

This paper proposes a method to generate synthetic hyperspectral images using the linear mixing model (LMM) but with a controllable synthetic variability and ground truth for both the endmembers and abundance maps. The method employs a generative adversarial network (GAN), specifically a conditional Wasserstein GAN (WGAN) with a gradient penalty, that learns to generate new samples from the learned data distribution of a real hyperspectral image (HSI). The diversity or variability of the sampled endmembers can be efficiently tuned by two parameters.

The rest of this paper is organized as follows. Section 2.1 briefly introduces GANs. Section 2.2 describes the proposed method. Section 3 presents the results of the experiments, and finally, conclusions are drawn in Section 4.

2. Materials and Methods
2.1. Generative Adversarial Networks

GANs were first introduced in 2014 [8] as a generative model that can learn the distribution \( P_r \) of the training data and generate new samples from the distribution. It consists of two neural networks, a generator \( G \), and a discriminator \( D \) that are trained simultaneously. The generator \( G \) takes a latent code \( z \) sampled from some prior distribution \( p(z) \) as an input and outputs a generated sample \( G(z) \). The discriminator \( D \) is trained to discriminate the generated samples from the real ones. \( G \) and \( D \) then compete with each other in a min-max game. The min-max objective of a GAN is

\[
\min_G \max_D V(G,D) = \mathbb{E}_{x \sim P_r} \left[ \log D(x) \right] + \mathbb{E}_{\tilde{x} \sim P_x} \left[ \log(1 - D(\tilde{x})) \right],
\]

where \( P_x \) is the model distribution, implicitly defined as \( \tilde{x} = G(z) \), \( z \sim p(z) \), and \( p \) is the distribution of the latent codes. The generator tries to minimize this objective function, and the discriminator tries to maximize it.

This objective function is not ideal as the generator saturates during training and stops learning as it gets too little feedback from the discriminator. To help stabilize training and prevent mode collapse in the generator, the Wasserstein GAN [15] with a gradient penalty was developed (WGAN-GP). The objective function becomes
\[
\min_{G} \max_{D \in \mathcal{F}} V(G, D) = \mathbb{E}_{x \sim P_r}[D(x)] - \mathbb{E}_{\tilde{x} \sim P_{\tilde{x}}}[D(\tilde{x})] + \lambda \mathbb{E}_{\hat{x} \sim \hat{P}_x}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2],
\]

where \( \mathcal{F} \) is the set of 1-Lipschitz functions \([16]\), \( \hat{x} \) is a convex combination of \( x \) from the real distribution and \( \tilde{x} \) from the model distribution, \( \hat{x} = cx + (1 - c)\tilde{x} \), \( c \) is a number between 0 and 1, and \( \lambda \) is the gradient penalty parameter. The gradient penalty softly enforces the critic (as the discriminator is called for WGANs) to be 1-Lipschitz, a necessary condition for the first two terms to be equivalent to the so-called Earth mover’s distance between the distributions. The critic outputs a nonnegative score instead of a probability.

2.2. The Proposed Method

The proposed method uses a conditional WGAN with gradient penalty and total variation \([17]\) (TV) regularization on the output of the generator. The TV regularization is needed to smooth the noisy output of the generator. A conditional GAN enables conditional generation by labeling the inputs to both the generator and discriminator. Thus, the latent codes are a combination of a noise vector sampled from a truncated Gaussian distribution and a label vector.

Figure 1a shows a schematic of the proposed conditional WGAN-GP-TV network. The pixels are labeled as Figure 1b illustrates. Since there were generally no class labels in the ground truth for the hyperspectral image used for training, we constructed them from the ground truth abundance maps for the image. Each pixel was labeled with a weighted one-hot vector, where the index of the one indicates the endmember class of the pixel as determined by which endmember has the greatest abundance fraction. The one-hot vector is then multiplied by the highest abundance which indicates the purity. The loss function used is

\[
\mathcal{L} = \mathbb{E}_{\hat{x} \sim \hat{P}_x, y \sim P_r(y)}[D(\hat{x}, y)] - \mathbb{E}_{x, y \sim P_r}[D(x, y)] + \lambda \mathbb{E}_{\hat{x} \sim \hat{P}_x, y \sim P_r(y)}[(\|\nabla_{\hat{x}} D(\hat{x}, y)\|_2 - 1)^2] + \lambda_{TV} \text{TV}(\tilde{x}).
\]

Here, \( y \) represents the labels we are conditioning on, \( \hat{x} \) is an interpolation between the fake and real samples, \( \lambda_{TV} \) is the total variation penalty parameter, and \( \text{TV}(\hat{x}) \) is the total variation of the generator’s output. It should be noted that the 2-norm of the gradient is generally used for the gradient penalty in the literature. Weight clipping can also be used to enforce the critic to be 1-Lipschitz, but this has been found to be unstable \([15]\). By conditioning both \( G \) and \( D \) on the labels, \( y \), the generator can generate specific endmembers with varying purity.

![Figure 1](image-url)

**Figure 1.** Schematics of the proposed conditional WGAN (a) and how the pixels are labeled (b). Both the generator and the critic are conditioned on the labels which indicate the class and purity of every pixel. (a) Schematic of the conditional WGAN-GP-TV. (b) The labeling of a pixel having an abundance vector \( s_p \). The vector \( y_p \) is the label.
To control the variability of the sampled endmembers, we employed the truncation trick [18], a latent sampling procedure where we sampled the codes $z$ from a truncated normal distribution. High truncation values correspond to a broad distribution, and low values correspond to a narrow distribution. Truncation trades fidelity for diversity. If we sample from a narrow distribution, then the variability of the sampled endmembers from the generator will be small, and vice versa. This can be interpreted as the extent to which the learned variability in the training data is used to generate new samples.

We constructed the synthetic HSIs by sampling a set of endmembers $A_p$ for every pixel $x_p$ having ground truth abundances $s_p$ and constructing the synthetic pixels as

$$x_p = A_p s_p. \quad (5)$$

The spectral variability is controlled by the truncation value and by specifying a minimum purity value. When sampling latent codes from the truncated normal, the purity is sampled uniformly between the minimum purity and 1.0. The ground truth endmembers for the synthetic HSI are obtained as the average of a large number of sampled endmembers.

### 3. Experiments and Results

#### 3.1. Experimental Set-Up

We trained the GAN on the widely used Urban, Samson, and Jasper datasets. The Urban HSI is 307 by 307 pixels and has 162 usable bands covering the 400–2500 nm wavelength range. We used the ground truth abundances for four reference endmembers. The Jasper HSI is 100 by 100 pixels with 198 usable bands and has 4 endmembers. The Samson HSI is 95 by 95 pixels, and the number of bands is 145, covering the 401–889 nm wavelength range. The dataset has the following endmembers: Water, Soil, and Tree. Figure 2 shows the simulated RGB images of the datasets.

![Figure 2](image)

**Figure 2.** Simulated and actual RGB images of the datasets used in the experiments. (a) Samson (simulated); (b) Urban (simulated); and (c) Apex (actual).

The GAN was trained for 300 epochs using the Adam optimizer for both the generator and critic. The critic was trained for five iterations for every iteration of the generator. The learning rate was set to 0.0002, the gradient penalty $\lambda$ was set to 10.0, and the TV penalty $\lambda_{TV}$ was set to 0.00003. The latent dimension of the sampled noise was set to 10.

#### 3.2. Evaluation

In order to investigate and compare the spectral variability of the original and synthetic images, we utilized a variational autoencoder. Specifically, the $\beta$-VAE model, as referred to in [19], was employed. The encoder of the VAE was used to generate images that illustrated the distributions of the encoding in a two-dimensional latent space.

Subsequently, we performed blind unmixing on the Urban synthetic datasets which demonstrated varying spectral variability. This was achieved using two distinct methods. The first method employed an autoencoder-based CNN-AEU as per [20], while the second involved a sparsity-constrained nonnegative matrix factorization termed $\ell_{1/2}$-NMF, as referenced in [21]. The $\ell_{1/2}$-NMF method separates the given HSI into two matrices: one...
of abundances and another of endmembers. This method was founded on the $\ell_{1/2}$-norm, a convex relaxation of the $\ell_0$-norm. The first method has been proven to be more robust against spectral variability as it incorporates the spectral-angle-distance (SAD) objective function, unlike the NMF method, which utilizes the MSE objective function. The NMF method is a traditional approach frequently used for unmixing and is appropriate for HSIs exhibiting low spectral variability.

We then compared our method with synthetic datasets constructed using both the linear mixing model (LMM) and the Hapke model as cited in [22] in order to introduce variability. The Hapke model is employed to simulate variability resulting from variations in topography. In this instance, we utilized it to generate several variants of reference endmembers that corresponded to different observation angles, assuming a Lambertian approximation.

To assess the performance of the unmixing methods, we employed the mean SAD between the estimated and ground truth endmembers:

$$mSAD = \frac{1}{R} \sum_{i=1}^{R} \arccos\left(\frac{\langle \hat{m}_i, m_i \rangle}{\|\hat{m}_i\|_2 \|m_i\|_2}\right),$$  

where $R$ is the number of endmembers, $\hat{m}_i$ represents the endmembers extracted by the method, and $m_i$ represents the reference endmembers. The lower the mSAD, the higher the similarity.

3.3. Results and Analysis

Figure 3 displays the original images, along with three synthetic images for each HSI with increasing spectral variability, in the VAE’s latent space. Figure 3d–f exhibit zero spectral variability, with fixed endmembers for every pixel (i.e., LMM reconstruction of the original image). As can be seen from Figure 3, the spectral variability for the high truncation values was akin to that in the original images, despite the simplified construction according to Equation (5).

Figure 4 demonstrates the effect of different truncation values on the grass endmember of the Urban dataset. The solid line represents the average, while the shaded region indicates the standard deviation. The plot used 150 samples from the GAN’s generator. It clearly shows that the variability increased with an increasing truncation value.

We performed blind unmixing on the synthetic Urban images displayed in the left column of Figure 3 and additionally on an image with a truncation of 0.4 and purity of 0.75. Figure 5 presents the results of the experiment in terms of the average SAD in radians from the ground truth endmembers. Each experiment comprised 10 runs. Figure 5 illustrates that the method using a scale-invariant objective function (CNNAEU) encountered no difficulties with increasing spectral variability, whereas the traditional method employing MSE exhibited progressively poorer performance. The $\ell_{1/2}$-NMF method recorded the lowest SAD for zero spectral variability. This fact, along with the low SAD values, suggests that the ground truth endmembers were a reliable ground truth for the synthetic images.

Figure 6 shows the latent codes for the original Urban dataset, a synthetic version created using the proposed method, and a synthetic version constructed using the Hapke model. As can be seen from Figure 6, the Hapke model did not capture the variability as well as the proposed method, suggesting that factors other than topography, such as natural biological variability in vegetation, were at work. Figure 7 shows the same encodings but for the Jasper dataset. Again, it can be seen that the Hapke model did not capture the variability as well as the proposed method. It should be noted that the latent code projection in these two figures is slightly different from that of Figure 3.
Figure 3. The latent space encodings of the original and synthetic images. The left column has the Urban images, and the right column has the Jasper images. Different colors indicate which endmember has the largest abundance fraction. The truncation and purity values are indicated in the titles and control the spectral variability of the synthetic images. (a) Original Urban HSI. (b) Original Jasper HSI. (c) Original Samson HSI. (d) Urban with zero variability. (e) Jasper with zero variability. (f) Samson with zero variability. (g) Urban with truncation value of 0.1 and purity of 0.8. (h) Jasper with truncation value of 0.1 and purity of 0.8. (i) Samson with truncation value of 0.1 and purity of 0.8. (j) Urban with truncation value of 0.1 and purity of 0.9. (k) Jasper with truncation value of 0.1 and purity of 0.9. (l) Samson with truncation value of 0.1 and purity of 0.9.
Figure 4. The figure shows the effect of increasing truncation on the grass endmember of the Urban dataset. The solid line is the average, and the shaded region indicates the standard deviation. The plot uses 150 samples from the generator of the GAN.

Figure 5. The average SAD from the ground truth endmembers for the original and three synthetic datasets with increasing spectral variability.
Figure 6. Latent space encodings of the original Urban image, the GAN simulated synthetic image using a truncation of 0.5 and purity of 0.8, and the Hapke model’s simulated synthetic image. (a) The original Urban image. (b) Image with GAN (truncation of 0.5 and purity of 0.8); and (c) using the Hapke model.

Figure 7. Latent space encodings of the original Jasper image, the GAN simulated synthetic image using a truncation of 0.5 and purity of 0.8, and the simulated image using the Hapke model. (a) The original Jasper image. (b) GAN with truncation of 0.5 and purity of 0.8. (c) Using the Hapke model.

We could further evaluate the method by comparing the mean and variance of the mostly pure pixels in the original and synthetic images. Figure 8 shows the average of mostly pure pixels (with the largest fractional abundance >0.8) for the original Urban HSI, the synthetic HSI with a truncation of 0.4 and purity of 0.75, and the Hapke synthetic HSI. As can be seen from the figure, the mean spectra for the original and GAN-simulated images were very similar in relative shape and scale and had incredibly similar variances (shaded areas). However, the magnitude of the mean spectra was higher for the image simulated using the proposed method. This was most likely a result of the LMM construction of the synthetic image. The mean spectra for the synthetic image generated using the Hapke model was not all that similar to the mean spectra for the original image.

Figure 9 shows simulated RGB images of the original Urban dataset and the synthetic images with a truncation of 0.4 and purity of 0.75. It is clear from the figure that there is a substantial difference between the original and synthetic images. The synthetic images appear to be much dimmer and flatter than the original image and also appear to be noisy. This is due to the simplistic LMM construction of the synthetic images and because the method is not spatial in nature (i.e., endmembers are randomly sampled for each pixel).
Figure 8. Average spectra of less mixed pixels (purity = 0.8) from the original Urban image, the GAN simulated synthetic image using a truncation of 0.4 and purity of 0.8, and the Hapke model’s simulated synthetic image. (a) The original Urban image. (b) GAN with truncation of 0.4 and purity of 0.8. (c) Using the Hapke model.

Figure 9. Simulated RGB images of the original Urban HSI and an HSI created using the proposed method. The method in [23] was used to create the images. (a) The original Urban image. (b) GAN with truncation of 0.4 and purity of 0.8.

4. Conclusions

In this paper, we proposed a technique that generates synthetic datasets that have a controllable degree of variability while having the ground truths for both the abundances and endmembers. The variability is generated by a conditional Wasserstein GAN that has learned the distribution of the original real image. The spectral variability is controlled by two parameters, the truncation of the prior normal distribution and the purity. Experiments showed that the generated spectral variability approximated the original one. Blind unmixing experiments showed that the ground truth reference endmembers were a good ground truth for the synthetic images and illustrated the effects of increasing variability on the performance of the methods.

Author Contributions: Conceptualization, B.P.; methodology, B.P.; formal analysis, B.P.; validation, M.O.U., J.R.S. and B.P.; investigation, B.P.; writing—original draft preparation, B.P.; writing—review and editing, J.R.S. and M.O.U.; visualization, B.P.; supervision, J.R.S.; project administration, J.R.S.; funding acquisition, J.R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Icelandic Research Fund under Grant 207233-051.

Data Availability Statement: Not applicable.
Conflicts of Interest: The authors declare no conflict of interest.

References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.