# MORGAN: Meta-Learning-based Few-Shot Open-Set Recognition via Generative Adversarial Network

Debabrata Pal[1, 2], Shirsha Bose[3], Biplab Banerjee[1], Yogananda Jeppu[2]
[1]Indian Institute of Technology, Bombay, [2]Honeywell Technology Solutions, India,
[3]Technical University of Munich, Germany
{deba.iitbcsre19, shirshabosecs, getbiplab, yvjeppu}@gmail.com

## Abstract

*In few-shot open-set recognition (FSOSR) for hyperspectral images (HSI), one major challenge arises due to the simultaneous presence of spectrally fine-grained known classes and outliers. Prior research on generative FSOSR cannot handle such a situation due to their inability to approximate the open space prudently. To address this issue, we propose a method, Meta-learning-based Open-set Recognition via Generative Adversarial Network (MORGAN), that can learn a finer separation between the closed and the open spaces. MORGAN seeks to generate class-conditioned adversarial samples for both the closed and open spaces in the few-shot regime using two GANs by judiciously tuning noise variance while ensuring discriminability using a novel Anti-Overlap Latent (AOL) regularizer. Adversarial samples from low noise variance amplify known class data density, and we use samples from high noise variance to augment "known-unknowns". A first-order episodic strategy is adapted to ensure stability in the GAN training. Finally, we introduce a combination of metric losses which push these augmented "known-unknowns" or outliers to disperse in the open space while condensing known class distributions. Extensive experiments on four benchmark HSI datasets indicate that MORGAN achieves state-of-the-art FSOSR performance consistently.* [1]

## 1. Introduction

Hyperspectral imaging (HSI) sensors capture the material reflectance from the earth's surface in a densely sampled wide range of wavelengths with a multitude of real-life applications. In the case of traditional closed-set HSI classification, a pixel is assigned to one of the available *known* classes based on its spectral-spatial properties. However, HSI datasets inherently put a caveat on their limited training data availability for certain *known* classes compared to its wide range of spectral bands referred to as the 'curse of dimensionality' [4]. Recent advancements in few-shot learning (FSL) [35, 7, 23] have become a de-facto standard in many computer vision applications to bridge this gap. FSL has also been explored successfully for HSI classification [25] over machine learning, and deep-learning-based approaches [18, 28]. Nevertheless, a closed-set trained model is prone to encountering outliers during testing in real-life scenarios due to i) being deployed in a new geographical area ii) the presence of unannotated *known* class pixels due to the cost involved in HSI labeling. It becomes evident that a reliable HSI classifier should also be aware of outliers, apart from its pristine *known* class distribution knowledge. As a result, the emergence of FSOSR for HSI datasets has garnered increasing interest in the geoscience community [1, 21, 26] (Fig. 1).

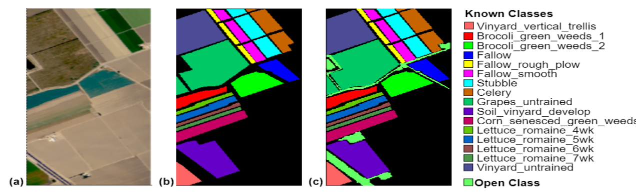


Figure 1. (a) The original image of the Salinas HSI dataset is shown. (b) The annotated land cover maps of *known* classes are displayed as per the actual ground truth. (c) One more open class is further annotated with existing *known* classes, which often get erroneously recognized as one of the *known* classes.

Existing OSR [2, 8, 11] and FSOSR [13, 20] methods tend to apply an empirical threshold on the model predictions to demarcate the outliers from the *known* classes. Utilizing such a threshold is seemingly discouraging as it is a purely dataset-dependent approach. Thanks to the recently introduced Outlier Calibration Network (OCN) [26], a three-layer binary classifier that meta-learns the pseudo-decision boundary between the *known* and outliers is found to improve the FSOSR performance. Interestingly, when we delved deeper into [26], we found $\approx 10\%$ lower open accuracy than closed-set for the 5-shot evaluation on the Indian Pines dataset, i.e, severe misclassification was observed for

[1] `https://github.com/DebabrataPal7/MORGAN`

open class 'Grass-pasture-mowed' which is spectrally similar to *known* 'Grass-pasture' and 'Grass-trees' classes.

It is evident that [26] cannot handle the scenario where fine-grained classes may simultaneously be present in the closed and open set. We suspect this is due to the limited generalization ability of [26] since it meta-learns a closed-set boundary from the available *base* classes without paying attention to approximate the broad knowledge of the open space. This invariably affects the open-set classification performance in HSI. For fine separation of the open and closed spaces, we argue that the presence of representative open samples of varied similarity measures with the closed-set data is necessary. This leads to our first research question *how to hallucinate fine-grained open space samples from the available closed-set data for better learning the known-unknown class separator in FSOSR?*

The generative modality hallucination techniques using GANs [27] so far follow a fully supervised approach and generate in-distribution data. Naturally, they cannot be trained optimally in the few-shot regime, as estimating the data density from a few training samples is extremely difficult. Additionally, they follow a closed-set model and cannot perform open space data synthesis. This motivates us to ask the next research question on *how to stably train a GAN model to simultaneously synthesize closed-set and open-set samples from few available closed-set training samples?*

Finally, we seek to obtain a dense and discriminative feature space for the closed-set samples while ensuring that the generated open space samples are scattered over a region. This will ensure the model to estimate the open space distribution better. In this regard, [26] ensures the discriminative closed-set representation; however, the open space scattering is overlooked in [26] as it does not explicitly model the open space distribution. This brings in the third research direction of *ensuring a discriminative closed-set distribution and diversity of the open space concurrently in FSOSR.*

***Our contributions***: To tackle the aforesaid problems, we propose a novel dual-conditional-GAN-based generative model called MORGAN, where two different noise variance values are used to hallucinate the *"pseudo-known"* and *"known-unknown"* samples. We propose to augment the closed-set data using *"pseudo-known"* features generated from an abating Gaussian distribution with a low noise variance in MORGAN. In parallel, we train a *"known-unknown"* generator from high noise variance and the hallucinated outliers are coupled with annotated *known-unknown* queries to maximize the open-world diversities. Then, a novel AOL regularizer asserts the discriminability between these *pseudo-known* features and outliers by restricting the *known-unknown* generator to synthesize samples from non-overlapping regions of the low and high noise variance components. To stabilize GAN training from a few samples, we proposed to integrate episodic first-order optimization-based meta-learning, Reptile [23], in our MORGAN framework for simultaneous pseudo-closed and open sample generation. Learning to reject these fine-grained class-specific outliers tighten the *known* class prototypes and put constraints on *known* class distribution to span inside a closed boundary in the metric space. Also, instead of only increasing population density per *known* class, we hypothesize that open space risk is further minimized by scattering out fine-grained boundary outliers to an open space. In order to execute the same, we introduce Outlier Scattering Loss to maximize the separation between the fine-grained pseudo-outliers and the closed-set distributions. Chiefly, our contributions are summarized as follows:

- To amplify data density in FSOSR, we develop a meta-learning-based method, MORGAN, which simultaneously generates class-conditioned pseudo *known* features and outliers by simulated controlling low and high noise variance. Also, a new regularizer, AOL, is proposed to distinguish pseudo outliers from the generated *pseudo-known* features.
- We adapt episodic first-order optimization-based meta-learning to stabilize GAN. Besides, we propose a prototype-based Outlier Scattering Loss to maximize the separation between closed-set boundary and open space and incorporate other metric losses to optimize the feature extractor.
- Proposed MORGAN is a lightweight model and gets optimized faster, which we have shown in qualitative analysis.

We conduct extensive experiments by assigning a few spectrally similar classes as open-closed pairs and thereby learn to reject fine-grained outliers on benchmark HSI datasets [2], namely, Indian Pines, Pavia, Salinas, and Houston-2013.

## 2. Related works

**Few-shot open-set recognition:** The problem of FSOSR is attempted in PEELER [20] by forming Gaussian clusters with a limited support set. Even though the closed-set distribution is learned well, outliers are not pushed away from the prototypes, failing to reject spectrally fine-grained outliers. SnaTCHer [13] thresholds the difference between original prototypes and query replaced transformed prototypes to find outliers. Contemporary FSOSR methods on HSI [1, 21] apply a threshold to reject outliers. However, it is hard to find an optimal threshold to reject outliers with a marginal spectral difference to the *known* land-cover samples. Again, reciprocal points classification loss [3] lag compact *known* class distribution in FSL, causing lower closed-set accuracy. Hence, we hypothesize that adversarial feature augmentation is required to simultaneously boost *known* class distribution and open-set in the FSOSR context.

**Generative open-set recognition:** A fine-grained open-set classifier must know the enriched closed-set distribution and be well up on the open space. Existing generative OSR

[2] http://www.ehu.eus/ccwintco/index.php/Hyperspectral Remote Sensing Scenes

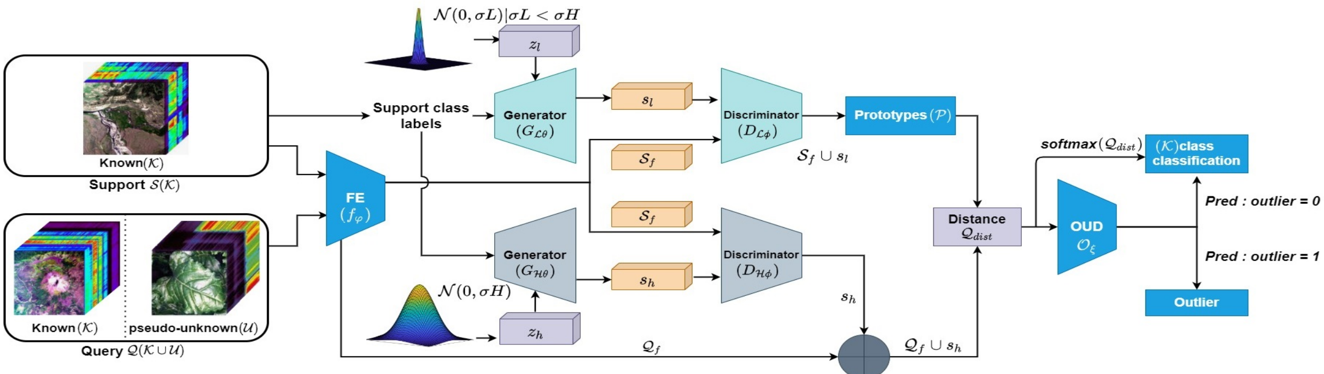



Figure 2. The framework of the proposed MORGAN. First, Feature extractor ($f_\varphi$) produces support features $\mathcal{S}_f$ for the *known* classes and query features $\mathcal{Q}_f$ for a combination of *known* and *pseudo-unknown* classes. Then, an adversarial process optimizes a latent vector sampled from isotropic Gaussian to produce *pseudo-known* features $s_l$ by Generator ($G_{\mathcal{L}\theta}$). $s_l$ is augmented with $\mathcal{S}_f$ to enrich the closed-set distribution, and prototypes $\mathcal{P}$ are computed. To enrich open space, we augment query features with $s_h$ obtained from Generator ($G_{\mathcal{H}\theta}$) by sampling another Gaussian noise. Outlier detector ($\mathcal{O}_\xi$) classifies a query from augmented query set $\mathcal{Q}_{aug}$ as an outlier based on its Euclidean distance $\mathcal{Q}_{dist}$ from $\mathcal{P}$. For *known* query prediction, we obtain its class by applying softmax over $\mathcal{Q}_{dist}$.

methods disjointly prioritize 1) reconstruction error-based outlier rejection [5, 33, 29] 2) pseudo-open-set generation [8, 22], and 3) closed-set distribution enrichment [31, 37]. Majorly generative OSR methods apply GAN [16] or autoencoder [24] utilizing large-scale training datasets. Class conditional auto-encoder (C2AE) [24] suffers from *known* sample selection bias from the training set for obtaining low reconstruction error on the *known* samples. OpenGAN [16] augments fake features and penalizes discriminator to learn to reject outliers. We strongly argue that all the adversarial samples are not always outliers; they can also be *known* class representative samples. The fundamental challenge in generative-FSOSR is the non-convergence of GAN from limited training data [6]. The adopted training strategy of MORGAN seems to be of help in this regard.

**Few-shot feature hallucination:** DAWSON [19] generate only *pseudo-known* samples by quickly adapting a new domain using optimization-based meta-learning, namely MAML [7] and Reptile [23] but it can not generate out-of-distributions data. MetaGAN [38] generates only fake data for few-shot classification and FAML [30] generates fake images by concatenating two noise vectors with a feature vector using unsupervised meta-learning. D$^2$GAN [17] introduces an anti-collapse regularizer to maintain discriminability and diversity in a few-shot regime. This regularizer minimizes the logarithmic similarity between generated adversarial samples to the logarithmic similarity of unit variance Gaussian noise. The strategy for fusing input images with random interpolation coefficients in F2GAN [12] can generate arbitrary landcover classes. FSGAN [32] magnifies singular values related to age, pose, skin tone from singular value decomposition [10] performed on StyleGAN2 [14]. However, the range of endmembers in the HSI datasets is infinite, leading to intractable singular values.

# 3. Proposed methodology

## 3.1. Preliminaries

FSOSR addresses the problem of open-set recognition with a few *known* training samples. Precisely, a meta-learner learns to reject the outliers together with the accurate classification of the *known* class samples using an episodic strategy. To this end, the meta-learner explores two disjoint sets of classes, *base* classes for meta-training and target classes for meta-testing. We select a random subset of *base* classes as *known-unknown* to enrich the open space, with the rest acting as *known* classes. For a given $\mathcal{K}$ *known* classes with $m$ training samples per class, we denote the support set as $\mathcal{S} = \{(x_i^s, y_i^s)\}_{i=1}^{m.\mathcal{K}}$ and is also termed as $\mathcal{K}$-way $m$-shot classification. Similarly, we form the query set as $\mathcal{Q} = \{(x_j^q, y_j^q)\}_{j=1}^{N.(\mathcal{K}\cup\mathcal{U})}$ with $N$ samples from each of the $\mathcal{K}$ *known* and $\mathcal{U}$ *known-unknown* classes. $x^s, x^q \in \mathbb{R}^{H\times W\times B}$ represent the 3D HSI patches of height $H$, width $W$, spectral bands $B$, and $y^s$, $y^q$ denotes the associated support and query set labels, respectively.

## 3.2. Overview of the MORGAN components

This section will first introduce the network architecture. Thereby, we explain the concept of a few-shot class-conditional closed and open feature generation strategy. Finally, we discuss the idea of the proposed AOL regularizer to reject the overlapping fine-grained pseudo-outliers.

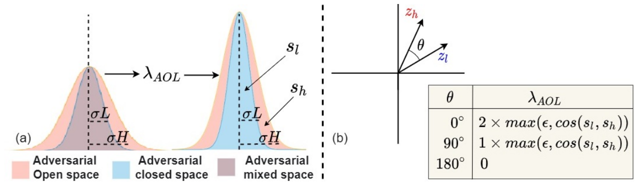
**A. Network architecture:** Fig. 2 illustrates the model architecture for MORGAN. Considering the 3D HSI patches as the inputs, we choose R3CBAM from OCN [26] as the backbone feature extractor ($f_\varphi$) due to its superior spectral-spatial HSI feature learning ability with the help of an attention-based CBAM3D layer. $f_\varphi$ produces 64D feature vector. We construct both MORGAN generators i.e., $G_{\mathcal{L}\theta}$,

$G_{\mathcal{H}\theta}$, using three consecutive dense layers with ReLU activation having 32, 48 and 64 units. During meta-training, we augment *pseudo-known* features $(s_l)$ produced by $G_{\mathcal{L}\theta}$ with support features $(\mathcal{S}_f)$ for better approximation of the closed-set distribution, where $\mathcal{S}_f$ is obtained by passing $\mathcal{S}$ through $f_\varphi$. The *Known* class prototypes $\mathcal{P}$ are measured by computing the class-wise mean of the augmented closed set features $\mathcal{S}_f(\mathcal{K}) \cup s_l$. Similarly, query features $(\mathcal{Q}_f)$ are augmented with fine-grained pseudo open samples $(s_h)$ by $G_{\mathcal{H}\theta}$ for better open space estimation. Discriminators $D_{\mathcal{L}\phi}$, $D_{\mathcal{H}\phi}$ have five dense layers, with ReLU activation having 48, 32, 16, 8 and 1 unit. The objective of $D_{\mathcal{L}\phi}$ is to classify whether the input feature is real $\mathcal{S}_f$ or *pseudo-known* $s_l$, whereas $D_{\mathcal{H}\phi}$ classifies the input feature as real $\mathcal{S}_f$ or pseudo-outlier $s_h$. We use a binary classifier with three dense layers having 16, 8 and 2 units as outlier detector $\mathcal{O}_\xi$. During meta-training, $\mathcal{O}_\xi$ learns to reject the outliers $\mathcal{Q}_f(\mathcal{U}) \cup s_h$. Finally, during meta-testing, an unlabeled query is first predicted by $\mathcal{O}_\xi$, as an outlier or *known* sample. In the case of *known* prediction, its class is estimated based on the distance from the prototypes $\mathcal{P}$

**B. Conditional open-closed feature generation in FSL:** In literature, a generator in D$^2$GAN [17] produces diverse closed-set samples utilizing two noise vectors. In contrast, we sample two noise vectors to generate pseudo closed and open samples employing dual-GANs individually responsible for closed-set and open space enrichment. However, the fundamental challenge in GAN meta-learning is obtaining sufficient gradients from limited training data. Again, GAN generates adversarial samples from an unknown probability distribution that is not optimized using a closed-form likelihood function. DAWSON [19] invokes optimization-based meta-learning to bridge GANs' likelihood-free training and obtain gradients in FSL paradigm. Inspired by [19], we also adapt first-order optimization-based meta-learning to train MORGAN dual-conditional GANs for the FSOSR task.

$G_{\mathcal{L}\theta}$ uses noise vector $z_l$ sampled from general isotropic Gaussian $[\mathcal{N}(0, \sigma L)]$. Due to low noise variation by $\sigma L$, $D_{\mathcal{L}\phi}$ penalizes $G_{\mathcal{L}\theta}$ to generate slightly fake *known* features $s_l$, such that augmenting $s_l$ with $\mathcal{S}_f$ boosts the closed-set representation. Similarly, $D_{\mathcal{H}\phi}$ penalizes $G_{\mathcal{H}\theta}$ that receives latent vector $z_h$ from another isotropic Gaussian distribution $\mathcal{N}(0, \sigma H)$ and produces highly fake feature $s_h$, where $\sigma H > \sigma L$. We consider $s_h$ as *pseudo-unknowns*. During meta-training, $\mathcal{O}_\xi$ learns to reject $s_h$ as outliers, and intuitively, it makes each *known* class distribution to span inside a closed boundary in metric space. Also, due to generating $s_l$ from Gaussian distribution, the probability of augmented closed-set samples reduces from its prototype following the CAP [34] theory. Thus a discriminative boundary of each *known* class distribution is established in FSOSR.

The training of two disjoint generators with $z_l \subset z_h$, can produce overlapping *pseudo-known* and pseudo-outlier




| $\theta$ | $\lambda_{AOL}$ |
|---|---|
| 0° | $2 \times max(\epsilon, cos(s_l, s_h))$ |
| 90° | $1 \times max(\epsilon, cos(s_l, s_h))$ |
| 180° | 0 |

Figure 3. (a) Sampling noise vectors from overlapping distribution causes mode collapse of the pseudo-open-closed features. AOL regularizer penalizes $G_{\mathcal{H}\theta}$ to generate $s_h$ from the noise vectors sampled from an isotropic Gaussian's mutually exclusive orange region as $s_l$ are generated from the blue region. (b) Effect of $z_l, z_h$ closeness in determining the ideal $\lambda_{AOL}$ limits.

samples. Essentially, $G_{\mathcal{H}\theta}$ should utilize the noise vectors sampled from $\{\mathcal{N}(0, \sigma H)\} - \{\mathcal{N}(0, \sigma L)\}$ region in Fig. 3a, or $s_h$ should be generated from noise vectors of standard deviation $(\sigma L < \sigma < \sigma H)$ of an isotropic Gaussian distribution. To make it mathematically amenable, we propose the AOL regularizer aiming to generate disjoint $s_l$, $s_h$.

**C. Anti-overlap latent regularizer:** To disentangle overlapping noise vectors that can produce adversarial outliers $s_h$ with equivalent feature representation to that of *pseudo-known* samples $s_l$, we define AOL regularizer. To the best of our knowledge, we are the first to introduce a novel regularizer in generative networks to control disentangled feature generation from overlapping noise variance. Specifically, we sample $z_l$, $z_h$ and generate synthesized features from the two generators for that class, i.e., $s_l = G_{\mathcal{L}\theta}(\mathcal{S}_f, z_l)$ and $s_h = G_{\mathcal{H}\theta}(\mathcal{S}_f, z_h)$. Since, $z_l \subset z_h$, some $s_h$ samples are anticipated to be collapsed to the same mode of $s_l$. To alleviate this, we define the AOL regularization term,

$$\lambda_{AOL}(z_l, z_h, s_l, s_h) = (1 + cos(z_l, z_h)).max(\epsilon, cos(s_l, s_h)) \tag{1}$$

Where $cos(i, j) = \frac{i.j}{\|i\|_2 \|j\|_2}$ represents the cosine similarity between vectors $i, j$ and $\epsilon$ is a small positive constant. We consider cosine similarity in formulating $\lambda_{AOL}$ due to its inherent bounding nature of prediction variance [9], causing empirical risk minimization. $\lambda_{AOL}$ penalizes $G_{\mathcal{H}\theta}$ for the similarity between adversarial features $s_l, s_h$ when their corresponding noise vectors $z_l, z_h$ become very similar. We minimize $\lambda_{AOL}$ in optimizing $G_{\mathcal{H}\theta}, D_{\mathcal{H}\phi}$ parameters in (2). The effect of this regularizer can be visualized in Fig. 3b. Without using it, $z_l, z_h$ sampled from overlapping distribution can reduce discriminability between $s_l, s_h$.

## 3.3. Learning and inference protocol

For each randomly sampled episode, we extract $\mathcal{S}_f$ and $\mathcal{Q}_f$ through feature extractor $f_\varphi$. Then, we optimize MORGAN generators $G_{\mathcal{L}\theta}, G_{\mathcal{H}\theta}$ and discriminators $D_{\mathcal{L}\phi}, D_{\mathcal{H}\phi}$ for loss functions $\mathcal{L}_l, \mathcal{L}_H$, respectively, in Algorithm 2. The generated adversarial samples are augmented to enrich the closed and open space. We compute summation of three-loss components $\mathcal{L}_{FE}$ considering i) *known*-class compaction loss $\mathcal{L}_{Kc}$, ii) outlier scattering loss $\mathcal{L}_{Os}$ and iii) outlier calibration loss $\mathcal{L}_{Oc}$ at the end of the episode. Finally, we optimize $\mathcal{O}_\xi$ for $\mathcal{L}_{Oc}$ and $f_\varphi$ for $\mathcal{L}_{FE}$. The steps for MORGAN meta-learning is shown in Algorithm 1.

**Meta-learning for conditional GANs:** For a few-shot classification task, Reptile [23] computes the model weight difference between its previously trained and latest trained versions using a newly sampled episode and performs the first-order optimization. Reptile quickly initializes task-specific parameters by generating good gradients from limited data. In MORGAN, first-order Reptile [23] is used in optimizing dual-conditional-GANs over second-order MAML [7].

$$\begin{aligned}\mathcal{L}_l &= \min_{G_{\mathcal{L}\theta}} \max_{D_{\mathcal{L}\phi}} \mathbb{E}_{s\sim\mathcal{S}_f}[log D_{\mathcal{L}\phi}(s|y^s)] \\ &+ \mathbb{E}_{z\sim\mathcal{N}(0,\sigma L)}[log(1-D_{\mathcal{L}\phi}(G_{\mathcal{L}\theta}(z|y^s)]; \\ \mathcal{L}_h &= \min_{G_{\mathcal{H}\theta}} \max_{D_{\mathcal{H}\phi}} \mathbb{E}_{s\sim\mathcal{S}_f}[log D_{\mathcal{H}\phi}(s|y^s)] \\ &+ \mathbb{E}_{z\sim\mathcal{N}(0,\sigma H)}[log(1-D_{\mathcal{H}\phi}(G_{\mathcal{H}\theta}(z|y^s)] + \lambda_{AOL}\end{aligned} \tag{2}$$

Where, $\mathcal{L}_h$, $\mathcal{L}_l$ are adversarial losses for generating $s_h$, $s_l$.

To optimize the MORGAN objective in (2), We initially update a clone copy of GAN model parameters using stochastic gradient descent in Algorithm 2. Based on input noise vector $z \in \{z_l, z_h\}$, the inner loop generates $s_l$ or $s_h$. We assign zero value for input $s$ and $z$ in optimizing $G_{\mathcal{L}\theta}$, $D_{\mathcal{L}\phi}$ and use $\mathbb{1} \in \{0,1\}$ as an indicator function which evaluates to 0 iff $[s = z = 0]$; thus $\lambda_{AOL}$ vanishes for $\mathcal{L}_l$ computation in (2). $\lambda_{AOL}(z, z', s, s')$ only penalizes $G_{\mathcal{H}\theta}$ and $D_{\mathcal{H}\phi}$ due to $\mathbb{1}$ is set as 1 when $[s \neq 0, z \neq 0]$. Finally, these gradients are returned to Algorithm 1 to update original parameters of discriminators and generators as,

$$\begin{aligned}&\mathcal{H}\phi \leftarrow \mathcal{H}\phi - \beta H \times \nabla DH; \mathcal{H}\theta \leftarrow \mathcal{H}\theta - \beta H \times \nabla GH \\ &\mathcal{L}\phi \leftarrow \mathcal{L}\phi - \beta L \times \nabla DL; \mathcal{L}\theta \leftarrow \mathcal{L}\theta - \beta L \times \nabla GL\end{aligned} \tag{3}$$

Where $\mathcal{L}\phi, \mathcal{L}\theta, \mathcal{H}\phi, \mathcal{H}\theta$ are the model parameters, $\beta L, \beta H$ are the learning rates and $\nabla DL, \nabla GL, \nabla DH, \nabla GH$ are the gradients obtained inside the inner loop.

**Estimating the prototypes:** We utilize the generated $s_l$ to enrich the closed-set distribution. Then, we compute the classwise mean of augmented support features to estimate individual *known* class prototypes $\mathcal{P}_k$. Also, to optimize for outlier scattering loss, we compute the open space prototype $\mathcal{P}_\mathcal{U}$ using the known-unknown queries.

$$\mathcal{P}_k = \frac{1}{m}\sum_{x_i^s \in \mathcal{S}^k} f_\varphi(x_i^s) \cup s_l; \mathcal{P}_\mathcal{U} = \frac{1}{N}\sum_{x_j^q \in \mathcal{Q}(U)} f_\varphi(x_j^q) \cup s_h; \tag{4}$$

Where $\mathcal{S}^k$ is the real support set of $k^{th}$ *known* class.

**Feature extractor and outlier detector optimization:** Our objective function to update $f_\varphi$ is an integral of threefold loss components focused on each *known* class distribution density maximization, fine-grained pseudo-outlier recognition, and scattering them from closed-set distribution. We augment *known-unknown* queries with pseudo-outliers to increase the open-set data density and compute loss for each query in $\mathcal{Q}_{aug}$, where, $\mathcal{Q}_{aug} \leftarrow \mathcal{Q}_f(\mathcal{U}) \cup s_h$.

**Algorithm 1: MORGAN meta-learning steps**

**Input:** $\mathcal{S}(\mathcal{K})$, $\mathcal{Q}(\mathcal{K}\cup\mathcal{U})$, $y^s$, $y^q$, $f_\varphi$, $\mathcal{O}_\xi$, $G_{\mathcal{L}\theta}$, $G_{\mathcal{H}\theta}$, $D_{\mathcal{L}\phi}$, $D_{\mathcal{H}\phi}$, $\sigma L$, $\sigma H$, $\beta L$, $\beta H$, iterations: $\mathcal{I}_o$, $\mathcal{I}_i$, InnerLoop learning rates: $\alpha L$, $\alpha H$

***Meta-Training Phase***:

1 Extract features: $\mathcal{S}_f \leftarrow f_\varphi(\mathcal{S}(\mathcal{K}))$, $\mathcal{Q}_f \leftarrow f_\varphi(\mathcal{Q}(\mathcal{K}\cup\mathcal{U}))$ ;

2 Train GANs: Randomly initialize $\mathcal{L}\theta$, $\mathcal{L}\phi$, $\mathcal{H}\theta$, $\mathcal{H}\phi$ and Sample episodes $\tau \sim\mathcal{S}(\mathcal{K})$ ;

**for** $i \leftarrow 1$ **to** $\mathcal{I}_o$ **do**

  **2.1** Clone parameters: $D_{\widehat{\mathcal{L}\phi}} \leftarrow D_{\mathcal{L}\phi}$, $D_{\widehat{\mathcal{H}\phi}} \leftarrow D_{\mathcal{H}\phi}$, $G_{\widehat{\mathcal{L}\theta}} \leftarrow G_{\mathcal{L}\theta}$, $G_{\widehat{\mathcal{H}\theta}} \leftarrow G_{\mathcal{H}\theta}$ ;

  **2.2** $\nabla DL, \nabla GL, s_l, z_l \leftarrow$***InnerLoop***$(D_{\widehat{\mathcal{L}\phi}}, G_{\widehat{\mathcal{L}\theta}}, D_{\mathcal{L}\phi}, G_{\mathcal{L}\theta}, \alpha L, \mathcal{I}_i, \tau, \sigma L, 0, 0)$;

  **2.3** $\nabla DH, \nabla GH, s_h, z_h \leftarrow$***InnerLoop***$(D_{\widehat{\mathcal{H}\phi}}, G_{\widehat{\mathcal{H}\theta}}, D_{\mathcal{H}\phi}, G_{\mathcal{H}\theta}, \alpha H, \mathcal{I}_i, \tau, \sigma H, s_l, z_l)$;

  **2.4** Update $\mathcal{L}\phi$, $\mathcal{L}\theta$, $\mathcal{H}\phi$, $\mathcal{H}\theta$ using $\beta L$, $\beta H$ (3);

3 Augment closed-set distribution and compute $\mathcal{P}$ (4);

4 Augment queries $\mathcal{Q}_{aug}$ for fine-grained open space;

5 Compute $\mathcal{L}_{FE}$ (8) using $\mathcal{L}_{Kc}$ (5), $\mathcal{L}_{Os}$ (6), $\mathcal{L}_{Oc}$ (7);

6 Optimize $\mathcal{O}_\xi$, $f_\varphi$ by $\mathcal{L}_{Oc}$, $\mathcal{L}_{FE}$, respectively.

**return** Updated parameters $\varphi$, $\mathcal{L}\theta$, $\mathcal{L}\phi$, $\mathcal{H}\theta$, $\mathcal{H}\phi$, $\xi$;

***Meta-Testing Phase***:

**Output:** Classification of the test-set query samples

7 $\mathcal{S}_f \leftarrow f_\varphi(\mathcal{S}(\mathcal{K}))$, $\mathcal{Q}_f \leftarrow f_\varphi(\mathcal{Q}(\mathcal{K}\cup\mathcal{U}))$ ;

8 Pass $\mathcal{S}_f$ to $G_{\mathcal{L}\theta}$, augment support and Compute $\mathcal{P}$;

9 Compute query distances: $\mathcal{Q}_{dist} \leftarrow \|\mathcal{P} - \mathcal{Q}_f\|_2^2$ ;

10 Classify $\mathcal{Q}_f$ samples of by passing $\mathcal{Q}_{dist}$ to $\mathcal{O}_\xi$ ;

**if** $\mathcal{O}_\xi$ *classifies query as known* **then**

  Predict class by applying softmax over $\mathcal{Q}_{dist}$ ;

**else**

  Classify the query as an outlier; ;

**return** Predicted class of the query samples;

**Algorithm 2: MORGAN Inner Loop training**

**Input:** $D_{\widehat{\phi}}$, $G_{\widehat{\theta}}$:, $D_\phi$, $G_\theta$, $\alpha$, $\mathcal{I}_i$,$\tau$, $\sigma$, $s$, $z$

**Output:** $\nabla D, \nabla G, s'$: new adversarial sample, $z'$

1 Randomly draw $K$ Samples $\{s_1, ..., s_K\} \sim q_{\tau_i}$ ;

2 **for** $i \leftarrow 1$ **to** $\mathcal{I}_i$ **do**

  **for** $k \leftarrow 1$ **to** $K$ **do**

    **2.1** $s' \leftarrow G_{\widehat{\theta}}(z')$ with $z' \sim \mathcal{N}(0,\sigma)$;

    **2.2** Update $D_{\widehat{\phi}}$ and $G_{\widehat{\theta}}$ parameters:

    $\widehat{\phi} \leftarrow \phi - \alpha \times \nabla_{\widehat{\phi}}(\mathcal{L}^D_{\widehat{\theta},\widehat{\phi}}(s_k, s') + \mathbb{1}.\lambda_{AOL})$;

    $\widehat{\theta} \leftarrow \theta - \alpha \times \nabla_{\widehat{\theta}}(\mathcal{L}^G_{\widehat{\theta},\widehat{\phi}}(s') + \mathbb{1}.\lambda_{AOL})$;

  **2.3** Compute: $\nabla D \leftarrow \phi - \widehat{\phi}$, $\nabla G \leftarrow \theta - \widehat{\theta}$ ;

**return** $\nabla D, \nabla G, s', z'$

**i.Known-class compaction loss:** Maintaining a compact abating representation of each *known* class requires minimizing the distance of each *known* query and support feature from the corresponding prototype. The optimization in (5) for each *known* query's euclidean distance minimizes intra-class variance of *known* classes and thereby helps in maximizing closed-set distribution density.

$$\mathcal{L}_{Kc} = \mathbb{E}_{q \in \mathcal{Q}_f(\mathcal{K}) \cup \mathcal{S}_f \cup s_l} \left[ -\log \frac{e^{-d(q,\mathcal{P}_y)}}{\sum_{k=1}^{\mathcal{K}} e^{-d(q,\mathcal{P}_k)}} \right] \quad (5)$$

Where, $\mathcal{P}_y$ is the true prototype of query $q$ and $d$ is the squared Euclidean distance.

**ii.Outlier scattering loss:** To scatter an outlier query, $\in \mathcal{Q}_{aug}$ from a *known* class distribution, the metric distance of that query from the *known* class true prototype $\mathcal{P}_y$ should be maximized. Effectively, it minimizes the probability of that query belonging to that *known* class. However, in multi-class classification, only repelling outliers from a set of *known* classes creates an uncertain residence of the outliers in metric space. Hence, we pull these outliers towards an open space prototype $\mathcal{P}_{\mathcal{U}}$ and gain transferable knowledge to scatter over the episodes. Also, penalizing the non-parametric prototype helps a set of outliers collectively and quickly move towards an open space to maximize the separation margin. Thus, we minimize the risk of misclassifying a fine-grained outlier to a *known* instance.

$$\mathcal{L}_{Os} = \mathbb{E}_{q \in \mathcal{Q}_{aug}} \left[ -\log \frac{e^{\gamma d(q,\mathcal{P}_y)}}{\sum_{k=1}^{\mathcal{K}} e^{\gamma d(q,\mathcal{P}_k)} + e^{-d(q,\mathcal{P}_{\mathcal{U}})}} \right] \quad (6)$$

Where, $\gamma$ is a positive repel factor to control the distance of an outlier query $\in \mathcal{Q}_{aug}$ from $\mathcal{P}_y$.

**iii.Outlier calibration loss:** We pass the Euclidean distance ($\mathcal{Q}_{dist}$) of each query from the prototypes to $\mathcal{O}_\xi$. In each episode, $\mathcal{O}_\xi$ learns a transferrable knowledge of classifying a query $\in \mathcal{Q}_f(\mathcal{K})$ as *known* sample and a query $\in \mathcal{Q}_{aug}$ as an outlier. Although $\mathcal{S}_f, s_l$ could be classified as *known*, we experimentally found no significant performance improvement by including $\mathcal{S}_f, s_l$. Using cross-entropy loss, threshold-free $\mathcal{O}_\xi$ parameters are optimized,

$$\mathcal{L}_{Oc} = \mathbb{E}_{q \in \mathcal{Q}_f(\mathcal{K}) \cup \mathcal{Q}_f(\mathcal{U}) \cup s_h} \left[ -\sum_{i=1}^{2} t_i log(\mathcal{O}_\xi(\mathcal{Q}_{dist})_i) \right] \quad (7)$$

Where, $t_i \in \{0, 1\}$ is the label for *known* or outlier class.

**Total feature extractor loss:** Finally, we compute the overall loss function to optimize $f_\varphi$ parameters in (8).

$$\mathcal{L}_{FE} = \mathcal{L}_{Kc} + \mathcal{L}_{Os} + \mathcal{L}_{Oc} \quad (8)$$

**Inference strategy:** We measure $\mathcal{Q}_{dist}$ of a target query from the prototypes. $\mathcal{O}_\xi$ classifies it as *known* or outlier based on input $\mathcal{Q}_{dist}$. In case of *known* class prediction, its class is further obtained by applying softmax over $\mathcal{Q}_{dist}$.

# 4. Experiments

## 4.1. Datasets and preprocessing

We evaluate MORGAN on four benchmark HSI datasets. Using AVIRIS sensor, **Indian Pines (IP)** dataset was acquired in northwestern Indiana. IP has $145 \times 145$ pixels with 220 bands covering 16 land-cover classes. **Salinas** captured at Salinas Valley, California, covers 16 labeled classes. It has $512 \times 217$ spatial dimension with 204 spectral bands. **University of Pavia** dataset was captured using ROSIS sensor for nine land cover classes having $610 \times 610$ pixels with 103 bands. We consider one more Salinas and six Pavia classes annotated in [21] for FSOSR. The **Houston-2013** dataset has $349 \times 1905$ pixels with 144 bands and was captured at the University of Houston for 15 landcover classes.

We apply PCA [36] on each of the IP, Pavia, and Salinas datasets to reduce spectral dimensionality to 30 bands and ten bands for the Houston dataset preserving 99% data variance for an individual dataset. Then, we slice cubic patches of dimension (11, 11, $ch$) at each pixel location using zero padding, where $ch$ indicates the number of spectral bands.

## 4.2. Evaluation metrics

To evaluate MORGAN performance, we use the standard OSR metrics, namely Closed Overall Accuracy (ClosedOA), Open Overall Accuracy (OpenOA), and AUROC (Area Under Receiver Operating Characteristics Curve). ClosedOA indicates the percentage of *known* class samples correctly classified. OpenOA evaluates the model in presence of outliers in (9), and AUROC refers to the outlier detection capability under various threshold configurations.

$$OpenOA = \frac{\sum_{k=1}^{\mathcal{K}+1} TP_k + TN_k}{\sum_{k=1}^{\mathcal{K}+1} TP_k + TN_k + FP_k + FN_k} \quad (9)$$

Where, $TP_k, TN_k, FP_k, FN_k$ represent the true positive, true negative, false positive, and false negative of the $k^{th}$ *known* class, respectively. All the outliers are combinedly considered as $\mathcal{K} + 1^{th}$ class in *OpenOA*.

## 4.3. Experimental protocol

We pick ten random classes as *base* classes for each dataset during meta-training and assign the remaining for meta-testing. Again, we split the *base* classes into spectrally fine-grained open-closed class pairs. Then, we form a query set with 15 samples per class from six randomly chosen *base* classes. Also, the support set is formed with *m* disjoint samples from each of the three random query classes. We use Adam optimizer [15] with a learning rate of 0.0001 for $\beta L$, $\beta H$ in Algorithm 1 and a value of 0.003 for $\alpha L$, $\alpha H$ in Algorithm 2. We set $\gamma$ as 1 in (6), $\epsilon$ as 0.00001 in (1) and $\sigma L$, $\sigma H$ as 0.2 and 1.0 (refer Table 4). During testing, we sample random episodes 500 times from target classes as per literature and record the mean $\pm$ standard deviation results in Table 1 and 2 for reliable prediction. All the methods are compared using the same experimental settings.

Table 1. 1-shot FSOSR performance comparison of the proposed MORGAN and SOTA Methods on the hyperspectral datasets

| Model | Indian Pines | | | Pavia University | | | Salinas | | | Houston-2013 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ClosedOA** | **OpenOA** | **AUROC** | **ClosedOA** | **OpenOA** | **AUROC** | **ClosedOA** | **OpenOA** | **AUROC** | **ClosedOA** | **OpenOA** | **AUROC** |
| OpenMax [2] | 43.33±0.63 | 48.54±0.29 | 44.44±0.33 | 52.08±0.53 | 54.61±0.21 | 52.22±0.35 | 51.92±0.52 | 42.50±0.23 | 50.33±0.34 | 37.50±0.73 | 31.33±0.71 | 36.11±0.33 |
| RDOSR [1] | 51.28±0.34 | 50.13±0.41 | 47.29±0.23 | 50.85±0.27 | 51.68±0.32 | 55.45±0.42 | 59.14±0.35 | 60.19±0.33 | 54.23±0.21 | 58.92±0.27 | 63.53±0.51 | 61.08±0.34 |
| MDL4OW [21] | 46.15±0.21 | 46.50±0.23 | 48.66±0.32 | 56.66±0.22 | 54.89±0.21 | 51.11±0.34 | 58.33±0.21 | 57.28±0.31 | 52.77±0.32 | 41.66±0.43 | 43.42±0.34 | 43.88±0.32 |
| PEELER [20] | 71.41±0.31 | 75.45±0.24 | 71.84±0.21 | 57.18±0.35 | 71.55±0.31 | 52.39±0.18 | 65.14±0.53 | 69.63±0.43 | 57.78±0.33 | 46.95±0.41 | 74.34±0.17 | 53.75±0.32 |
| SnaTCHer [13] | 89.33±0.11 | 81.25±0.23 | 74.53±0.32 | 58.50±0.29 | 75.57±0.34 | 50.35±0.33 | 65.94±0.67 | 78.91±0.28 | 53.88±0.43 | 58.73±0.35 | 77.42±0.33 | 48.05±0.34 |
| OCN [26] | 80.67±0.63 | 82.72±0.34 | 75.93±0.32 | 59.50±0.21 | 84.77±0.34 | 89.96±0.43 | 82.64±0.31 | 74.27±0.32 | 81.36±0.21 | 60.84±0.32 | 78.93±0.46 | 74.26±0.17 |
| MORGAN [Ours] | **91.47±0.14** | **87.42±0.08** | **90.83±0.12** | **79.88±0.16** | **85.22±0.21** | **90.11±0.09** | **83.24±0.17** | **90.22±0.14** | **91.96±0.12** | **77.15±0.22** | **89.20±0.15** | **91.06±0.12** |

Table 2. 5-shot FSOSR performance comparison of the proposed MORGAN and SOTA Methods on the hyperspectral datasets

| Model | Indian Pines | | | Pavia University | | | Salinas | | | Houston-2013 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ClosedOA** | **OpenOA** | **AUROC** | **ClosedOA** | **OpenOA** | **AUROC** | **ClosedOA** | **OpenOA** | **AUROC** | **ClosedOA** | **OpenOA** | **AUROC** |
| OpenMax [2] | 51.92±0.34 | 58.33±0.51 | 55.62±0.46 | 69.17±0.32 | 58.12±0.53 | 53.36±0.53 | 69.23±0.25 | 58.12±0.35 | 55.44±0.34 | 41.66±0.24 | 35.64±0.35 | 37.84±0.51 |
| RDOSR [1] | 55.98±0.51 | 55.92±0.45 | 52.38±0.52 | 64.74±0.45 | 64.89±0.34 | 63.94±0.37 | 67.80±0.35 | 66.32±0.51 | 60.28±0.47 | 68.82±0.52 | 66.37±0.31 | 62.56±0.43 |
| MDL4OW [21] | 50.76±0.53 | 46.96±0.35 | 64.51±0.51 | 65.33±0.25 | 62.66±0.19 | 48.66±0.41 | 75.01±0.37 | 62.66±0.34 | 72.88±0.22 | 46.66±0.21 | 44.42±0.23 | 46.22±0.33 |
| PEELER [20] | 82.81±0.39 | 87.37±0.34 | 74.19±0.25 | 60.71±0.23 | 72.69±0.53 | 60.36±0.51 | 74.20±0.52 | 75.39±0.31 | 60.39±0.36 | 57.98±0.47 | 75.57±0.57 | 55.47±0.34 |
| SnaTCHer [13] | 92.00±0.51 | 89.42±0.45 | 76.05±0.55 | 74.91±0.35 | 78.10±0.38 | 54.98±0.29 | 74.21±0.54 | 83.15±0.39 | 72.11±0.43 | 69.24±0.34 | 83.29±0.52 | 49.09±0.45 |
| OCN [26] | 94.61±0.38 | 84.71±0.51 | 88.40±0.38 | 71.55±0.32 | 87.94±0.34 | 91.90±0.45 | 84.88±0.37 | 85.61±0.39 | 88.52±0.34 | 71.97±0.29 | 84.96±0.48 | 88.50±0.47 |
| MORGAN [Ours] | **95.09±0.18** | **90.43±0.24** | **95.59±0.12** | **81.11±0.19** | **92.18±0.18** | **92.98±0.14** | **86.64±0.09** | **93.70±0.11** | **94.99±0.11** | **83.64±0.21** | **92.18±0.14** | **95.17±0.16** |

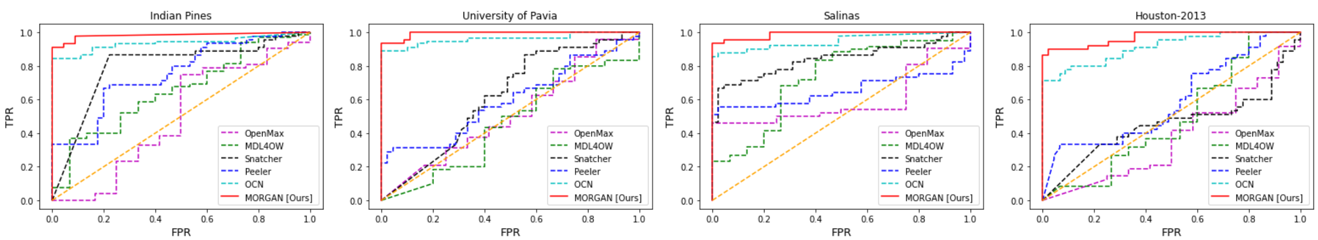



Figure 4. 5-shot AUROC comparison using different FSOSR methods on four benchmark HSI datasets. MORGAN (indicated in 'Red' color) shows the best True-Positive Rate (TPR) performance for a low False-Positive Rate (FPR) value over the other methods.

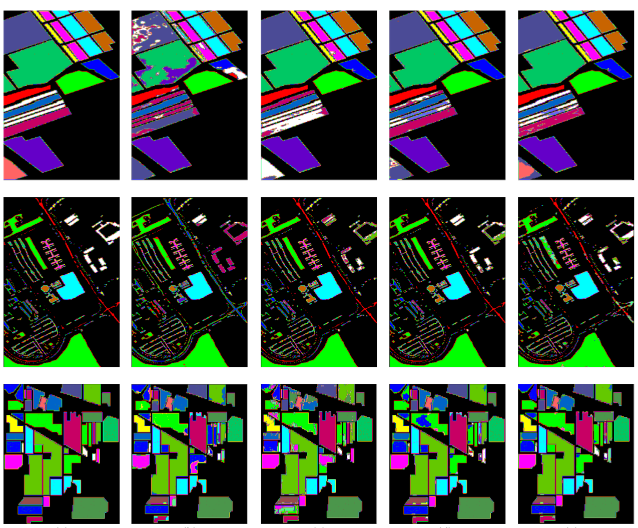



Figure 5. Comparison of 5-shot FSOSR classification maps by SOTA methods, namely b) PEELER (c) SnaTCHer (d) OCN, and proposed (e) MORGAN over (Top) Salinas (Middle) University of Pavia and (Bottom) Indian Pines. The ground truth is shown in (a) for each dataset with the open classes annotated in 'White' color.

## 4.4. Experimental results

We compare MORGAN against the state-of-the-art (SOTA) 1-shot and 5-shot FSOSR methods in Table 1 and 2, respectively. Thanks to MORGAN's fine-grain outlier separation ability; it shows better 1-shot and 5-shot FSOSR performance than other SOTA methods over HSI datasets.

OpenMax [2] and MDL4OW [21], initially developed for large-scale OSR, are adapted in FSOSR by fitting Weibull distribution in Prototypical Networks[35]. RDOSR [1] performs OSR over HSI datasets in latent space using limited supervised samples. PEELER [20], SnaTCHer [13], and OCN [26] were developed for the FSOSR context and readily evaluated on HSI datasets. For 1-shot FSOSR, we see MORGAN beating the next best alternative by 4.7% OpenOA, 14.9% AUROC over IP, a significant 20.38% ClosedOA over Pavia, 11.31% OpenOA, 10.6% AUROC over Salinas and 16.31% ClosedOA, 10.27% OpenOA, 16.8% AUROC raise over the Houston-2013 dataset. The 5-shot FSOSR performance also appears quite promising by MORGAN. Over IP, it boosts 7.19% AUROC over other methods.A stellar performance, 8.09% OpenOA gain over Salinas, 11.67% ClosedOA, and 7.22% OpenOA hike are observed over the Houston-2013 dataset. Overall, MORGAN achieves the maximum area under the ROC curve in Fig. 4 reflecting high fine-grained outlier recognition capability in a few-shot context. The classification maps are compared in Fig. 5. PEELER and SnaTCHer misclassify a few *known* classes for Salinas and Pavia. MORGAN recognizes fine-grained open class 'Grass-pasture-mowed' in IP, 'Building 2' in Pavia better than other methods. The closed-set recognition performance is also superior in MORGAN. Like 'Vinyard_vertical_trellis' in Salinas, most of the *known* class samples MORGAN recognized correctly.

Table 3. Ablation study on our various novelties to optimize MORGAN's generative components for 5-shot FSOSR over four HSI datasets

| Model | Indian Pines | | | Pavia University | | | Salinas | | | Houston-2013 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ClosedOA | OpenOA | AUROC | ClosedOA | OpenOA | AUROC | ClosedOA | OpenOA | AUROC | ClosedOA | OpenOA | AUROC |
| Without $\mathcal{L}_{Os}$ | 88.31±0.37 | 83.39±0.29 | 80.32±0.45 | 80.22±0.28 | 77.54±0.37 | 82.43±0.22 | 81.75±0.28 | 86.55±0.31 | 91.79±0.30 | 81.48±0.29 | 86.57±0.31 | 90.34±0.25 |
| Without AOL | 94.57±0.39 | 86.71±0.25 | 89.45±0.37 | 79.34±0.33 | 84.28±0.27 | 88.26±0.23 | 83.58±0.34 | 87.31±0.44 | 88.23±0.26 | 80.11±0.28 | 87.23±0.26 | 88.33±0.35 |
| Without GAN | 80.33±0.26 | 81.27±0.23 | 80.55±0.30 | 71.78±0.24 | 86.85±0.35 | 84.37±0.35 | 74.80±0.22 | 82.88±0.20 | 86.71±0.36 | 74.50±0.37 | 82.19±0.42 | 86.71±0.22 |
| Inner loop - MAML[7] | 90.71±0.43 | 84.60±0.24 | 88.29±0.22 | 80.59±0.25 | 81.96±0.39 | 74.73±0.27 | 84.35±0.28 | 86.85±0.21 | **95.81±0.10** | 81.98±0.45 | 88.69±0.35 | 89.65±0.27 |
| MORGAN [Ours] | **95.09±0.18** | **90.43±0.24** | **95.59±0.12** | **81.11±0.19** | **92.18±0.18** | **92.98±0.14** | **86.64±0.09** | **93.70±0.11** | 94.99±0.11 | **83.64±0.21** | **92.18±0.14** | **95.17±0.16** |

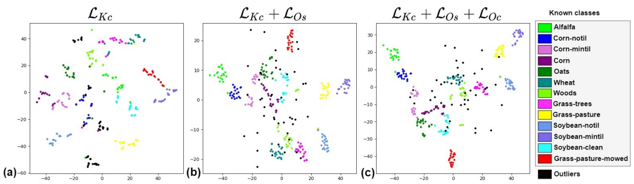



Figure 6. t-SNE visualization of the metric space due to optimizing MORGAN for various loss components over the IP dataset.

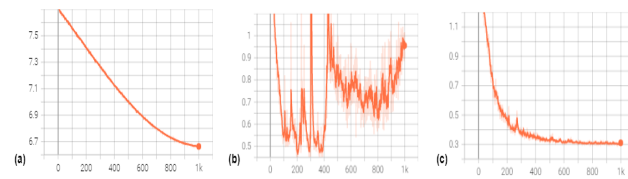


Figure 7. 5-shot FSOSR loss decay comparison over Salinas using (a) OCN [26], MORGAN with (b) MAML [7], and (c) Reptile[23].

### 4.5. Further analyis

**Effect of different loss components:** In Table 3, we observe that MORGAN gains 7.04% OpenOA for IP, 14.64% OpenOA, and, 10.55% AUROC for Pavia, 7.15% OpenOA for Salinas, incorporating outlier scattering loss. Fig. 6 shows the t-SNE plots incorporating different loss functions to train MORGAN. While optimizing by only compaction loss, a couple of fine-grained outliers fall inside the closed set distribution in Fig. 6a. Incorporating scattering loss creates a compact decision boundary separating the *known* and fine-grained outlier pair, in Fig. 6b. Further outlier calibration loss helps distinguish outliers better in Fig. 6c.

**Influence of AOL regularizer:** The AOL regularizer generates fine-grained outliers encompassing the *known* class distributions. It helps in boosting 3.72%, 7.9%, 6.39% OpenOA and 6.14%,4.72%,7.76% AUROC respectively, for IP, Pavia, and Salinas datasets in Table 3.

**Impact of feature generation:** We evaluated MORGAN without generating any adversarial *known* and pseudo-outliers. However, we observed a 14.76% ClosedOA, 9.16% OpenOA, and 15.04% AUROC drop for IP compared to our adversarial feature augmentation strategy in Table 3.

**Space complexity:** MORGAN is quite a lightweight model. Generators $G_{\mathcal{L}\theta}$, $G_{\mathcal{H}\theta}$ have only 5520 and discriminators $D_{\mathcal{L}\phi}$, $D_{\mathcal{H}\phi}$ have 6129 parameters individually. Feature extractor $(f_{\varphi})$ is common for all comparing methods with 38,114 parameters, and $\mathcal{O}_{\xi}$ has 218 parameters.

**Time complexity:** The dual-GANs in MORGAN are optimized by the Reptile-based first-order meta-learning [23] in Algorithm 2, producing the fastest loss convergence in Fig.7c. OCN [26] shows quite a long time to converge, and MAML [7]-based MORGAN suffers from fluctuation due to second-order optimization. Utilizing Reptile, accuracy also boosts by 10.22% OpenOA, a massive 17.45% AUROC for Pavia, and 5.5% AUROC for Houston dataset as shown in Table 3. However, MAML is slightly better for the Salinas dataset by a marginal 0.82%AUROC rise.

Table 4. Ablation study on the noise variance $(\sigma L, \sigma H)$ values of $G_{\mathcal{L}\theta}, G_{\mathcal{H}\theta}$ over the Indian Pines dataset for 5-shot FSOSR task

| Noise variance | | Indian Pines | | |
|---|---|---|---|---|
| $\sigma$L | $\sigma$H | ClosedOA | OpenOA | AUROC |
| **0.1** | **1.0** | 95.40±0.31 | 91.97±0.21 | 94.68±0.29 |
| **0.2** | **1.0** | **95.09±0.18** | **90.43±0.24** | **95.59±0.12** |
| **0.3** | **1.0** | 95.13±0.33 | 88.42±0.29 | 90.07±0.11 |
| **0.5** | **1.0** | 95.91±0.36 | 84.63±0.24 | 85.51±0.32 |
| **0.5** | **0.8** | 93.69±0.43 | 87.28±0.38 | 88.21±0.27 |
| **0.5** | **0.6** | 91.71±0.35 | 91.42±0.41 | 91.91±0.38 |
| **0.3** | **1.2** | 95.11±0.61 | 87.07±0.57 | 89.82±0.25 |

**Optimal noise variance values:** Table 4 shows the ablation study by varying the noise variance of two generators responsible for generating *pseudo-known* and outlier samples. There exists three scenarios, (1) increasing $\sigma L$ from 0.1 to 0.5 and keeping $\sigma H = 1.0$ (constant), OpenOA and AUROC steadily decrease as pseudo-outliers become false positives to the closed-set distribution. (2) keeping $\sigma L = 0.5$ (constant) and decreasing $\sigma H$ from 1.0 to 0.6, ClosedOA decreases, and OpenOA increases as many *pseudo-known* samples fall just outside its *known* class boundary. Finally, (3) increasing $\sigma H$ beyond 1.0 makes generated pseudo-outliers mix with other *known* class samples, reducing OpenOA. We find an optimal selection of $\sigma L$ = 0.2 and $\sigma H$ = 1.0 as the best noise variance combination for the two MORGAN generators in evaluating HSI datasets.

## 5. Conclusions

This paper proposes a novel FSOSR method, MORGAN, which simultaneously generates *pseudo-known* and outlier samples to enrich the closed and open space, respectively. Further, we introduce a new regularizer to retain distinguishability between adversarial *known*-outlier pairs. Also, we optimize the MORGAN feature extractor with three-fold loss functions to jointly squeeze closed-set distribution and spread away fine-grained outliers. Experimental results over four benchmark HSI datasets and ablation studies prove our superior performance over existing FSOSR methodologies. We plan to evaluate MORGAN for fine-grained HSI cross datasets experiments in the future.

## References


[1] Razieh Kaviani Baghbaderani, Ying Qu, Hairong Qi, and Craig Stutts. Representative-discriminative learning for open-set land cover classification of satellite imagery. In *European Conference on Computer Vision*, pages 1–17. Springer, 2020.

[2] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016.

[3] Guangyao Chen, Limeng Qiao, Yemin Shi, Peixi Peng, Jia Li, Tiejun Huang, Shiliang Pu, and Yonghong Tian. Learning open set network with discriminative reciprocal points. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 507–522. Springer, 2020.

[4] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 7(6):2094–2107, 2014.

[5] Lucas Deecke, Robert Vandermeulen, Lukas Ruff, Stephan Mandt, and Marius Kloft. Image anomaly detection with generative adversarial networks. In *Joint european conference on machine learning and knowledge discovery in databases*, pages 3–17. Springer, 2018.

[6] Chen Ding, Yu Li, Yue Wen, Mengmeng Zheng, Lei Zhang, Wei Wei, and Yanning Zhang. Boosting few-shot hyperspectral image classification using pseudo-label learning. *Remote Sensing*, 13(17):3539, 2021.

[7] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017.

[8] ZongYuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. Generative openmax for multi-class open set classification. *arXiv preprint arXiv:1707.07418*, 2017.

[9] Spyros Gidaris and Nikos Komodakis. Dynamic few-shot visual learning without forgetting. pages 4367–4375, 2018.

[10] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. In *Linear algebra*, pages 134–151. Springer, 1971.

[11] Mehadi Hassen and Philip K Chan. Learning a neural-network-based representation for open set recognition. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, pages 154–162. SIAM, 2020.

[12] Yan Hong, Li Niu, Jianfu Zhang, Weijie Zhao, Chen Fu, and Liqing Zhang. F2gan: Fusing-and-filling gan for few-shot image generation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2535–2543, 2020.

[13] Minki Jeong, Seokeon Choi, and Changick Kim. Few-shot open-set recognition by transformation consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12566–12575, 2021.

[14] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.

[15] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. iclr. 2015. *arXiv preprint arXiv:1412.6980*, 9, 2015.

[16] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. *arXiv preprint arXiv:2104.02939*, 2021.

[17] Kai Li, Yulun Zhang, Kunpeng Li, and Yun Fu. $D^2$GAN: A few-shot learning approach with diverse and discriminative feature synthesis. 2019.

[18] Wei Li, Chen Chen, Mengmeng Zhang, Hengchao Li, and Qian Du. Data augmentation for hyperspectral image classification with deep cnn. *IEEE Geoscience and Remote Sensing Letters*, 16(4):593–597, 2018.

[19] Weixin Liang, Zixuan Liu, and Can Liu. Dawson: A domain adaptive few shot generation framework. *arXiv preprint arXiv:2001.00576*, 2020.

[20] Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos. Few-shot open-set recognition using meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, 2020.

[21] Shengjie Liu, Qian Shi, and Liangpei Zhang. Few-shot hyperspectral image classification with unknown classes using multitask deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):5085–5102, 2020.

[22] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 613–628, 2018.

[23] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.

[24] Poojan Oza and Vishal M Patel. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2307–2316, 2019.

[25] Debabrata Pal, Valay Bundele, Biplab Banerjee, and Yogananda Jeppu. Spn: Stable prototypical network for few-shot learning-based hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 2021.

[26] Debabrata Pal, Valay Bundele, Renuka Sharma, Biplab Banerjee, and Yogananda Jeppu. Few-shot open-set recognition of hyperspectral images with outlier calibration network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3801–3810, January 2022.

[27] Shivam Pande, Avinandan Banerjee, Saurabh Kumar, Biplab Banerjee, and Subhasis Chaudhuri. An adversarial approach to discriminative modality distillation for remote sensing image classification. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0, 2019.

[28] Mercedes E Paoletti, Juan Mario Haut, Ruben Fernandez-Beltran, Javier Plaza, Antonio Plaza, Jun Li, and Filiberto

Pla. Capsule networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4):2145–2160, 2018.

[29] Pramuditha Perera, Vlad I Morariu, Rajiv Jain, Varun Manjunatha, Curtis Wigington, Vicente Ordonez, and Vishal M Patel. Generative-discriminative feature representations for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11814–11823, 2020.

[30] Aniwat Phaphuangwittayakul, Yi Guo, and Fangli Ying. Fast adaptive meta-learning for few-shot image generation. *IEEE Transactions on Multimedia*, 24:2205–2217, 2021.

[31] Stanislav Pidhorskyi, Ranya Almohsen, Donald A Adjeroh, and Gianfranco Doretto. Generative probabilistic novelty detection with adversarial autoencoders. *arXiv preprint arXiv:1807.02588*, 2018.

[32] Esther Robb, Wen-Sheng Chu, Abhishek Kumar, and Jia-Bin Huang. Few-shot adaptation of generative adversarial networks. *arXiv preprint arXiv:2010.11943*, 2020.

[33] Mohammad Sabokrou, Mohammad Khalooei, Mahmood Fathy, and Ehsan Adeli. Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3379–3388, 2018.

[34] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2317–2324, 2014.

[35] Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. *arXiv preprint arXiv:1703.05175*, 2017.

[36] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[37] Houssam Zenati, Manon Romain, Chuan-Sheng Foo, Bruno Lecouat, and Vijay Chandrasekhar. Adversarially learned anomaly detection. In *2018 IEEE International conference on data mining (ICDM)*, pages 727–736. IEEE, 2018.

[38] Ruixiang Zhang, Tong Che, Zoubin Ghahramani, Yoshua Bengio, and Yangqiu Song. Metagan: An adversarial approach to few-shot learning. *Advances in neural information processing systems*, 31, 2018.