Active Learning for Improved Semi-Supervised Semantic Segmentation in Satellite Images

Shasvat Desai *
Orbital Insight
shasvat.desai@orbitalinsight.com

Debasmita Ghose *
Yale University
debasmita.ghose@yale.edu

Abstract

Remote sensing data is crucial for applications ranging from monitoring forest fires and deforestation to tracking urbanization. Most of these tasks require dense pixel-level annotations for the model to parse visual information from limited labeled data available for these satellite images. Due to the dearth of high-quality labeled training data in this domain, there is a need to focus on semi-supervised techniques. These techniques generate pseudo-labels from a small set of labeled examples which are used to augment the labeled training set. This makes it necessary to have a highly representative and diverse labeled training set. Therefore, we propose to use an active learning-based sampling strategy to select a highly representative set of labeled training data. We demonstrate our proposed method's effectiveness on two existing semantic segmentation datasets containing satellite images: UC Merced Land Use Classification Dataset and DeepGlobe Land Cover Classification Dataset. We report a 27% improvement in mIoU with as little as 2% labeled data using active learning sampling strategies over randomly sampling the small set of labeled training data.

1. Introduction

Semantic segmentation has found vast applications in the domain of remote sensing, including but not limited to environmental monitoring [60, 62], land use classification, and change detection [11, 27, 4, 12, 18, 51]. The largest barrier to applying these segmentation techniques is the availability of representative labeled data across different geographies and terrains. Each pixel in a satellite image can represent a large area on the ground, thus requiring domain knowledge and experience to annotate pixel-level labels. This makes it significantly expensive in terms of cost and time to collect a large set of pixel-wise labels [23]. To alleviate this problem, recent work in the computer vision community has explored
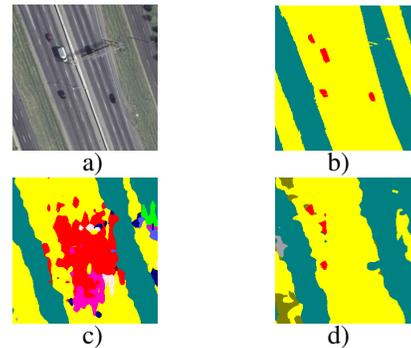

Figure 1: a) An image from the UC Merced Land Use Classification Dataset [59], b) Ground Truth of the same image provided by the DLSRD dataset [46], c) Baseline semi-supervised semantic segmentation model trained with 2% labeled data, d) Output of our active learning based semi-supervised semantic segmentation model trained with 2% labeled data.

using fewer pixel-wise labels along with information from unlabeled images in a semi-supervised fashion [17, 31, 48]. However, these small sets of images that are labeled pixel-wise are chosen randomly from a dataset [17, 31]. This might bias the semi-supervised network towards a particular set of classes, degrading its performance. Therefore, we propose to use active learning to select a representative set of labeled examples for semi-supervised semantic segmentation for land cover classification.

This work is the first to explore a semi-supervised approach to semantic segmentation in satellite images to the best of our knowledge. We use a conditional GAN [30] based on Mittal et al. [31] which takes in a small number of labeled examples and a large unlabeled pool of data. This conditional GAN generates pseudo-labels based on limited labeled examples to augment the labeled pool. This makes it essential to have a diverse set of labeled training data. Thus, we propose to use active learning to select a highly representative set of labeled training samples.

*Authors contributed equally

Active learning aims to select the most informative and representative data instances for labeling from an unlabeled data pool based on some information measure. We sample a subset of the images and their corresponding labels at random from a dataset, which serves as our labeled training set for the conditional GAN. We do the sampling again using an active learning-based sampling strategy which would provide a more diverse set of training data and show a performance improvement even when only very few training samples are available. With as little as 2% labeled data, we report an improvement of up to 27% in mIoU over random sampling. We demonstrate our proposed method's efficacy on two existing semantic segmentation datasets containing satellite images: UC Merced Land Use Classification Dataset [46, 59], and DeepGlobe Land Cover Classification Dataset [9].

Active learning for semantic segmentation [28, 56] yields patches of the given input image that are most informative. However, in this work, we require an active learning-based sampling strategy that gives us the set of most informative images from the given dataset. To achieve this, we propose using active learning for image classification to select entire images from the given dataset, which are the most informative. We then query and obtain dense-pixel level annotations only for the actively selected samples, giving us our diverse labeled training data for semi-supervised semantic segmentation.

Finally, we propose this method for sample selection to act as a guiding process for large-scale dataset creation, requiring the collection of dense pixel-level annotations. It would require significantly less cost and effort to obtain coarse image-level labels for the images and then use our proposed methodology to sample informative images labeled at pixel-level using image-classification-based active learning. The code has been made publicly available[*].

Our key contributions are summarized as follows:

- We use pool-based active learning sampling strategies to intelligently select labeled examples and improve performance for a GAN-based semi-supervised semantic segmentation network for satellite images.

- We demonstrate the applicability of the proposed method for selecting an optimal subset of data instances for which pixel-level annotations should be obtained.

2. Background and Related Work

2.1. Active Learning

Active Learning is a technique that uses a learning algorithm which learns to select samples from an unlabeled pool of data for which the labels should be queried.

---

[*]https://github.com/immuno121/ALS4GAN

Scenarios for Active Learning: Active Learning is typically employed in the following settings: [43]. *Membership Query Synthesis* [44] is a setting where the learner generates an instance from an underlying distribution. *Stream-based Selective Sampling* [1] queries each unlabeled instance one at a time based on some information measure. The last scenario *Pool-based sampling*, used in this paper, assumes a large pool of unlabeled data and draws instances from the pool according to some information measure.

Query Strategies: There are several query strategies in the pool-based sampling strategies to select samples for which we need to query labels.

The *margin sampling strategy* [2, 39] selects the instance that has the smallest difference between the first and second most probable labels.

$$x_M = \arg\max_x [P_\theta(\hat{y_2}|x) - P_\theta(\hat{y_1}|x)] \qquad (1)$$

Intuitively, instances with small margins are more ambiguous and knowing the true label should help the model discriminate more effectively.

The third and the most common strategy is *Entropy-based sampling* [24, 25].

$$x_H^* = \arg\max_x - \sum_y P_\theta(y|x) \log(P_\theta(y|x)) \qquad (2)$$

where y ranges over all possible labels of x. Entropy is a measure of a variable's average information content. So intuitively, this method selects samples by ranking them based on their information content.

Applications: Active learning techniques have found numerous applications in medical imaging [16, 19, 29, 45, 47] and remote sensing [20, 26, 38, 50, 54, 55] communities because obtaining labeled data in those domains have been particularly challenging [35, 42]. Some recent work have used deep learning techniques for active learning-based image classification [37, 45], semantic segmentation [28, 56, 58] and object detection [40]. However, to the best of our knowledge, this is the first work that uses a deep active learning-based image classifier to select labeled examples for a semi-supervised semantic segmentation network.

2.2. Semi-Supervised Semantic Segmentation

Most of the existing semi-supervised semantic segmentation techniques either use consistency regularization strategies [7, 13, 32, 33, 34] or generative model to augment pseudo-labels to existing labeled data pool [17, 48] or some combination of the two [31].

Consistency Regularization-based strategies: The core idea in consistency regularization is that predictions for unlabeled data should be invariant to perturbations. Some recent work has pointed out the difficulty in performing consistency regularization for semi-supervised seman-

tic segmentation because it violates the cluster assumption [13, 34]. Some other work [6, 13, 21] use data augmentation techniques like CutMix [61] and ClassMix [33], which composite new images by mixing two original images. They hypothesize that this would enforce consistency over highly varied mixed samples while respecting the original images' semantic boundaries.

GAN-based strategies: Souly et al. [48] was the first work to perform semi-supervised semantic segmentation using a GAN. They employ the generator to generate realistic visual data that forced the discriminator to learn better features for more accurate pixel classification. However, these generated images were not sufficiently close to the real images since it is challenging to generate realistic-looking images from pixel-wise maps.

To overcome the drawback of poorly generated images, Hung et al. [17] propose a conditional GAN. The generator is a standard semantic segmentation network that takes in images and their ground truth maps. The discriminator is a fully convolutional network (FCN) that takes the ground truth, and the segmentation map predicted by the generator and aims to distinguish between the two. Thus, it is difficult for the discriminator to determine if the pixels belong to the real or the fake distribution by looking at one pixel at a time without context.

Mittal et al. [31] propose to replace the FCN-based discriminator with an image-wise discriminator that determines if the image belongs to the real or the fake distribution, which is a relatively easy task. Additionally, they propose to use a supervised multi-label classification branch [52] which decides on the classes present in the image and thus aids the segmentation network to make globally consistent decisions. During evaluation, they fuse the two branches to alleviate both low-level and high-level artifacts that often occur when working in a low-data regime. In this work, we use the s4GAN branch of the network presented by Mittal et al. [31] and propose to select a more optimal set of labeled examples to improve the performance of the network over a random selection of labeled examples.

3. Our Method

We propose to use active learning techniques to select a small informative subset of labeled data that would help the semi-supervised semantic segmentation model learn more effectively with a representative pool of labeled data. The proposed framework is demonstrated in Figure 2. It should be noted that for any given image, our method assumes the ability to gain access to its corresponding image and pixel-level annotations.

3.1. Active Learning for Image Classification

Algorithm 1 describes how active learning was used to select the most informative and diverse set of labeled sam-

---

Algorithm 1: Active Learning for Labeled Sample Selection

Input:
Labeled ratio R and unlabeled pool of data X_N
Output:
Informative samples and their image-level labels (X_L, Y_L^I)
Define:
$learner \leftarrow$ Neural network based image classifier
$oracle \leftarrow$ source of labels

1 Number of data points to sample: $X_{N_L} \leftarrow R * X_N$
2 Size of initial labeled pool for learner:
  $init\_size \leftarrow \lceil \alpha_{init} * X_{N_L} \rceil$
3 Initial labeled pool for active learner:(X_{init}, Y_{init})
4 Train the active learner: $learner(X_{init}, Y_{init})$
5 Unlabeled pool: $X_{pool} \leftarrow X_N - X_{init}$
6 $X_L \leftarrow \{\}, Y_L^I \leftarrow \{\}$
7 Number of samples to query in each iteration :
  $N_Q = \beta_Q * init\_size$
  while $n(X_L) \leq X_{N_L}$ **do**
  Query Step:
  Inference on unlabeled pool using $learner$:
8   $prediction\_scores = learner(X_{pool})$
  Select top-Q most informative instances (N_Q):
9   $X_Q = sampling\_strategy(prediction\_scores)$
10   $Y_Q = oracle(X_Q)$
11   $X_{pool} = X_{pool} - X_Q$
12   $X_L = X_L \bigcup X_Q$
13   $Y_L^I = Y_L^I \bigcup Y_Q$
  Teach Step:
  Retrain the $learner$ with updated labeled pool:
14   $learner(X_L, Y_L^I)$
15 **end**
16 **return** X_L, Y_L^I

---

ples for semi-supervised semantic segmentation. We use active image classification and sample images using pool-based sampling strategies [43] as described in Section 2.1. Algorithm 1 is used as an offline process to sample informative samples and it accepts two inputs: labeled ratio R, and an unlabeled pool of data, X_N. The labeled ratio R, determines the number of labeled samples used to train the semi-supervised model. The labeled ratio R, used in this paper for each dataset, can be found in Table 1.

Initialization: The active learner is initialized with $init\_size$ number of image-level labels (line 3), which is a function of the number of labeled samples to be returned. We define the parameter $\alpha_{init} \in (0, 1]$ (line 2) to control the size of the initial labeled pool of the active learner. α_{init} helps in determining the optimal size of the labeled pool that the active learner should be initialized with for every labeled
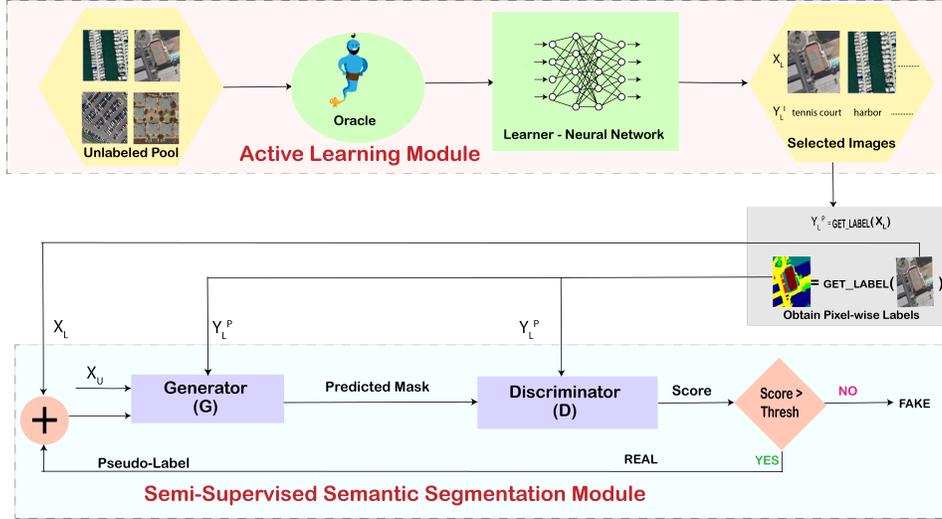
Figure 2: **Proposed Framework:** 1) The active learning module expects an unlabeled pool of data as its input. It is an image classification network that returns samples X_L selected based on some information measure determined by the sampling strategy used for the active learner. 2) A *get_label* operation is performed to obtain pixel-level labels corresponding to images returned by the active learning module. 3) A conditional GAN is then trained where the generator module is a semantic segmentation network and expects the labeled images returned by the active learning module, X_L, the corresponding pixel-wise labels for these samples Y_L^P, along with the remaining unlabeled images X_U. It outputs a segmentation mask. 4) The discriminator expects the predicted segmentation masks from the generator along with the pixel-wise ground truth labels Y_L^P, and outputs a prediction confidence score. 5) prediction masks with a score greater than the predefined confidence threshold τ are selected and treated as pseudo-labels to train the GAN and are augmented to the labeled pool as shown by the "+" sign on the bottom left corner.

ratio, R. A low value of α_{init} would result in the active learner being initialized with a tiny pool of data not providing sufficient information about the data distribution. In contrast, a large value of α_{init} might bias the active learner toward a particular set of initial samples, which might lead to under-sampling of a particular class. Intuitively, setting α_{init} to 1 would make to outcome close to being equivalent to random sampling. We found this approach to perform better than having a fixed initial labeled pool size, irrespective of the labeled ratio, R. The *learner* is then trained with this initial labeled pool (line 4). Once the initial labeled samples are selected for the active learner, they are removed from the unlabeled pool (line 5). We sample data instances and their labels by performing the query and teach steps in an interleaved fashion.

Query Step: We run inference using the trained *learner* on the entire unlabeled pool (line 8) and obtain the model's confidence scores for each sample in that pool. Then using some uncertainty measure based on the active learning strategy used, the oracle queries image-level labels for top-Q uncertain samples, N_Q. For instance, if entropy-based sampling [24, 25] is used, then the oracle will return labels for samples with the highest entropy measure. The optimal number of data instances N_Q, queried from the *oracle* in

every iteration is a function of the initial labeled pool size *init_size* and another parameter, $\beta_Q \in (0, 1]$ (line 7). We define β_Q to determine the number of iterations for which the active learner will be trained. It is crucial for the active learner's performance because a small value of β_Q will add only a small number of labels in each iteration, resulting in a negligible weight update of the active learner. In contrast, a large value of β_Q will cause a massive update in the learner's weights at every step. It will also reduce the total number of steps the learning algorithm will take to reach its target X_{N_L}. This will leave little room for the learner to learn from its mistakes in each iteration, directly impacting the quality labels produced. Once the labels for N_Q number of samples are queried from *oracle*, the images and their corresponding labels are added to the result set, X_L and Y_L^I (lines 12, 13).

Teach Step: In this step, the *learner* is trained with the updated labeled pool of samples obtained from the query step (line 14). The image classification network's capacity for the *learner* is also crucial in determining the quality of the selected samples. Any network with low capacity tends to underfit, while any network with a higher capacity than required could overfit and detrimentally affect the downstream task's performance.

Labeled Ratio(R)	2%	5%	12.5%	100%
UC Merced[59]	34	85	211	1680
DeepGlobe[9]	12	32	80	642

Table 1: Number of Labeled Examples per Labeled Ratio in the UC Merced and DeepGlobe datasets

3.2. Semi-Supervised Semantic Segmentation

We use the s4GAN network proposed by Mittal et al. [31] for performing semi-supervised semantic segmentation using a small number of pixel-wise labeled examples along with a pool of unlabeled examples. This is a conditional GAN-based technique where the generator G is a segmentation network. The generator takes in all the labeled and unlabeled images, along with the ground truth masks. The discriminator D takes the predicted segmentation map and the available ground truth masks concatenated with their respective images. The network attempts to match the real and the predicted segmentation maps' distribution through adversarial training.

Notation:
x_L^P, y_L^P: image with their pixel-wise labels
x_U^P: image with no pixel-wise ground-truth labels

3.2.1 Segmentation Network (Generator)

The segmentation network S is trained with loss L_S, which is a combination of three losses: the standard cross-entropy loss, the feature matching loss, and the self-training loss.
Cross Entropy Loss: Standard supervised pixel-wise cross entropy loss term evaluated only for the labeled samples x_L^p is shown is Equation 3.

$$L_{ce} = -\sum y_L^P \cdot log(S(x_L^P)) \qquad (3)$$

Feature-Matching Loss: The feature matching loss L_{fm} [41] aims to minimize the mean discrepancy between the feature statistics of the predicted, $S(x_U^P)$ and the ground truth segmentation maps, y_L^P as shown in Equation 4. This loss uses both labeled and unlabeled training examples.

$$L_{fm} = ||\mathbb{E}_{(y_L^P, x_L^P) \sim D_l} D(x_L^P \oplus y_L^P) -$$
$$\mathbb{E}_{(x_U^P) \sim D_u} D(x_U^P \oplus S(x_U^P))|| \quad (4)$$

Self-Training Loss: This loss is used for only unlabeled data. This loss aims to pick the best outputs of the segmentation network (i.e., those outputs that could fool the discriminator) that do not have a corresponding ground truth mask and reuse them for supervised training. Intuitively, it pushes the segmentation network to produce predictions

that the discriminator cannot distinguish from real. The discriminator's output is a score between 0 and 1, denoting the discriminator's confidence that the predicted segmentation mask is real. The predicted segmentation mask with a score greater than the predefined confidence threshold τ, is selected and treated as a pseudo-label to train the GAN.

Equation 5 describes the self-training loss.

$$L_{st} = \begin{cases} -\sum \mathbf{y*} \cdot log(S(x_U^P)) & \text{if } D(x_U^P) \geq \tau \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$\mathbf{y*}$ = pseudo pixel-wise labels which are the predictions of the segmentation network
Finally, the objective function for the generator is given by Equation 6.

$$L_S = L_{ce} + \lambda_{fm} L_{fm} + \lambda_{st} L_{st} \qquad (6)$$

where, $\lambda_{fm} > 0$ and $\lambda_{st} > 0$ are the weighting parameters for the feature matching and the self-training losses.

3.2.2 Discriminator

The discriminator is trained to distinguish between the real labeled examples and the fake segmentation masks generated by the network concatenated with the corresponding input images. It is trained using the original GAN loss proposed by Goodfellow et al. [14] as shown in Equation 7.

$$L_D = \mathbb{E}_{(y_L^P, x_L^P) \sim D_l} log D(x_L^P \oplus y_L^P) +$$
$$\mathbb{E}_{(x_U^P) \sim D_u} log(1 - D(x_U^P \oplus S(x_U^P)) \quad (7)$$

3.3. Labeled Example Selection for Semi-Supervised Semantic Segmentation using Active Learning

To obtain labeled samples for a semi-supervised semantic segmentation network, the proposed framework uses the active learning module from Algorithm 1 defined as $Active\_Sampler$ in Algorithm 2 and shown in Figure 2. The active learning module expects an unlabeled pool of data (X_N) and the labeled ratio R as its input. It returns samples X_{N_L} (line 4), which are selected based on some information measure determined by the sampling strategy used for the active learner. The active learning module is called only once to select informative data instances for the semi-supervised model. In the get_label stage, we obtain pixel-level labels corresponding to only those images returned by the $Active\_Sampler$ to serve as the initial labeled training data for the semi-supervised segmentation module. This enables the semi-supervised semantic segmentation network to learn with a representative labeled set instead of a random subset of the data.

The conditional GAN described in Section 3.2 is then trained, where the generator module is a semantic segmentation network and expects the labeled images returned by the active learning module X_{N_L}, the corresponding pixel-wise labels for these samples $Y_{N_L}^P$, along with the remaining unlabeled images, X_{N_U}. The output of the generator network is a segmentation mask. The discriminator expects the predicted masks along with the pixel-wise ground truth labels $Y_{N_L}^P$, and outputs a probability score between 0 and 1, denoting its confidence in the predicted mask being real and belonging to the ground truth. If this confidence score is greater than the predefined confidence threshold τ, then it implies the generator has successfully predicted a mask that appears real to the discriminator. Hence, this prediction is augmented to the ground-truth, $Y_{N_L}^P$ (line 14) and used as a pseudo-label, and the GAN is trained with this updated dataset. These pseudo-labels contribute to the self-training loss detailed in Section 3.2.1. The generator and the discriminator are trained adversarially until a stopping criterion is satisfied.

Algorithm 2: Semi-Supervised semantic segmentation using samples obtained from the Active Learner

---

Input:
Labeled Ratio, R and unlabeled pool, X_N
Define:
$Active\_Sampler \leftarrow$ Active Learning module
1   $\tau \leftarrow$ Confidence Threshold
2   $G \leftarrow$ Generator network of conditional GAN
3   $D \leftarrow$ Discriminator network of conditional GAN
4   Active sampler from Algorithm 1:
     $X_{N_L, \_} = Active\_Sampler(R, X_N)$
5   **Get_Label:** Obtain pixel-wise labels for the images returned by the active sampler:
     $Y_{N_L}^P = get\_label(X_{N_L})$
6   Images without pixel-wise labels:
     $X_{N_U} \leftarrow X_N - X_{N_L}$
   Semi-Supervised Semantic Segmentation:
   while $i <$ *iterations* **do**
7      $pseudo\_label = \{\}$
8   **for** $(x_{N_L}, x_{N_U}, y_{N_L}^P)$ *in* $(X_{N_L}, X_{N_U}, Y_{N_L}^P)$ **do**
9        $mask = G(x_{N_L}, x_{N_U}, y_{N_L}^P)$
10       $confidence = D(mask, y_{N_L}^P)$
11   **if** $confidence > \tau$ **then**
12         $pseudo\_label = pseudo\_label \bigcup mask$
13   **end**
14   Augment pseudo-labels to ground-truth and train the Generator model via self-training loss:
       $Y_{N_L}^P = Y_{N_L}^P \bigcup pseudo\_label$
15   **end**

---

4. Experiments and Results

4.1. Datasets and Evaluation Metric

UC Merced Land Use Classification Dataset: The UC Merced Land Use Classification dataset [59] has 2100 RGB images of size 256x256 pixels and 0.3m spatial resolution, with image-level annotations for each of the 21 classes. We use the pixel-level annotations for the UC Merced dataset made publicly available by Shao et al. [46] which has 17 classes as proposed in [3]. The dataset was randomly split into training and validation sets with 1680 training images (80%) and 420 validation images (20%).

DeepGlobe Land Cover Classification Dataset: The DeepGlobe land cover classification dataset is comprised of DigitalGlobe Vivid+ images of dimensions 2448x2448 pixels and spatial resolution of 0.5 m. There are 803 pixel-wise annotated training images, each with pixel-wise label covering seven land cover classes. Since there are no image-level annotations available for the DeepGlobe dataset, to generate image-level annotations, we calculate which class contains the highest number of pixels for every image and assign that particular coarse class to the image. The dataset was randomly split into training and validation sets with 642 training images (80%) and 161 validation images (20%).

Evaluation Metric: We use mean Intersection-over-Union (mIoU) as our evaluation metric.

4.2. Implementation Details

Active Learning for Image Classification We used ResNet 101 and ResNet 50 [15] as our image classification networks for the UC Merced and the DeepGlobe Datasets respectively which were trained using the Cross-Entropy loss. The network was trained using the SGD optimizer with a base learning rate of 0.001 and momentum of 0.9. We used a batch size of 4 and trained for 50 epochs after each query. Through cross-validation, we found the optimal value of $\alpha_{init} = 0.1$ and $\beta_Q = 0.5$. We implemented the network using the open-source skorch [53] framework. The different active learning query strategies were implemented using the modAL open-source toolbox [8] and trained on a single NVIDIA GTX-2080ti GPU.

Semi-Supervised Semantic Segmentation We use a GAN-based semi-supervised semantic segmentation technique proposed by Mittal et al. [31]. The generator is comprised of a segmentation network which in our case is DeepLabv2 [5] trained with a ResNet-101 [15] backbone pretrained on the ImageNet dataset [10]. The discriminator is a binary classifier with four convolutional layers with 4x4 kernels with 64, 128, 256, 512 channels each followed by a Leaky ReLU activation [57] with negative slope of 0.2 and a dropout layer [49] with dropout probability of 0.5. The segmentation network in the generator is trained with SGD optimizer base learning rate of 2.5e-4, momentum of 0.9,
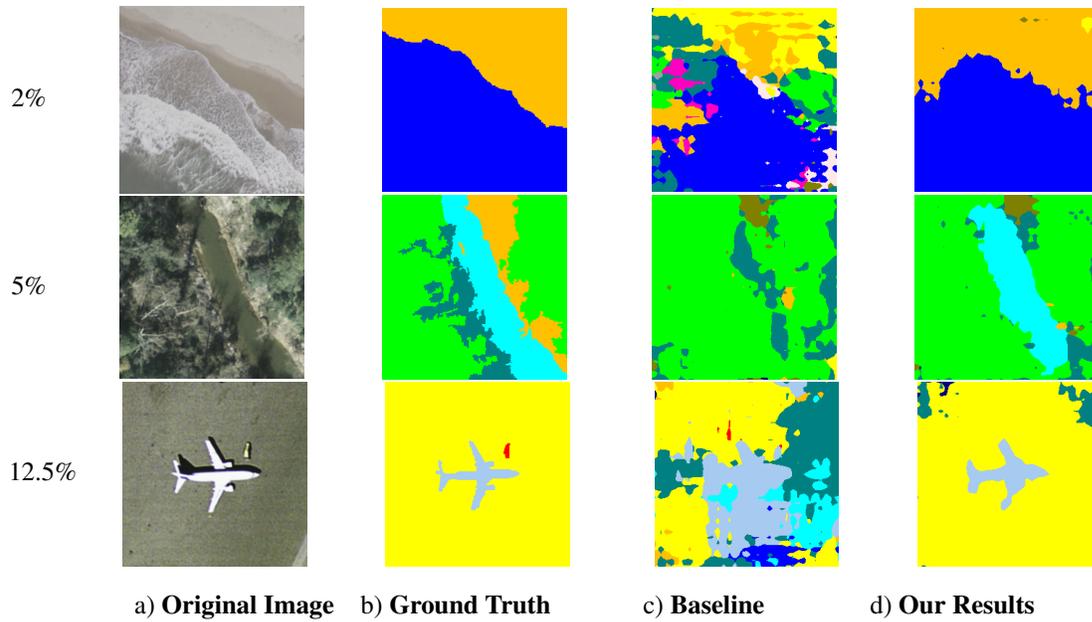
a) **Original Image**   b) **Ground Truth**   c) **Baseline**   d) **Our Results**

Figure 3: Qualitative Results from the UC Merced Land Use Classification Dataset for different labeled ratios

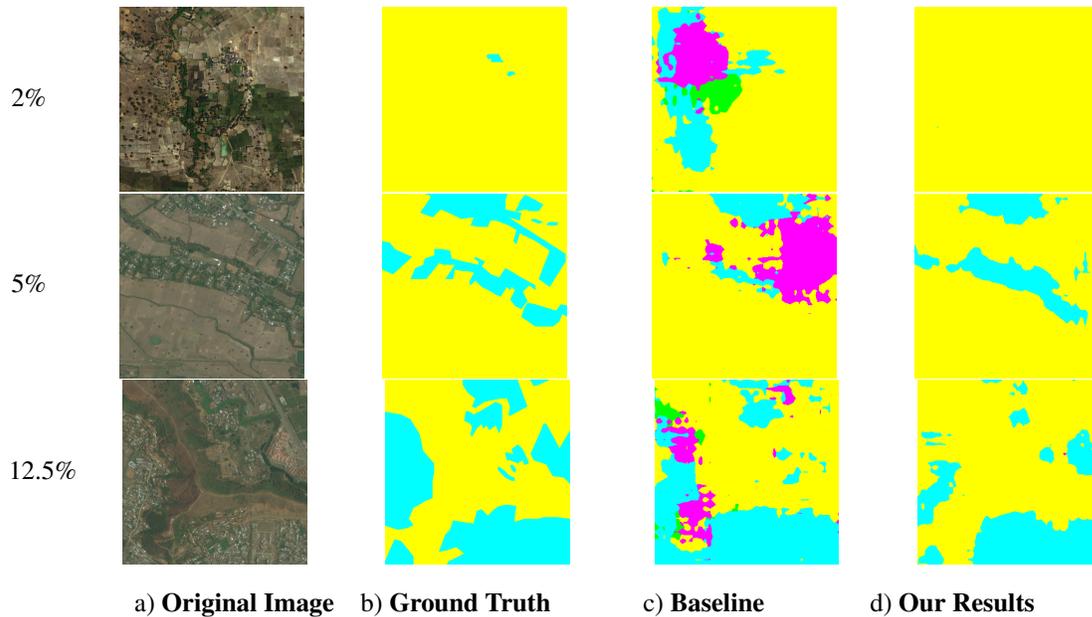a) **Original Image**   b) **Ground Truth**   c) **Baseline**   d) **Our Results**

Figure 4: Qualitative Results from the DeepGlobe Land Cover Classification Dataset for different labeled ratios

and a weight decay of 5e-4 as described in [17, 31]. The image classification network in the discriminator is trained using the Adam optimizer [22] with a base learning rate of 1e-4. Through cross-validation, we found the optimal loss weights to be $\lambda_{fm} = 0.1$ and $\lambda_{st} = 1.0$ and the optimal value of τ to be 0.6. For the DeepGlobe dataset, we resize each image to 320x320 pixels to reduce the training time.

We implemented the network using PyTorch [36] on four NVIDIA Tesla V100 GPUs.

4.3. Results and Analysis

Our baseline is a vanilla s4GAN [31] network, where the labeled data is selected randomly from a given dataset. We report the mean and standard deviation of mIoUs across

Labeled Ratio(R)	2%	5%	12.5%
s4GAN [31] (Baseline)	0.347 ± 0.0077	0.408 ± 0.0035	0.473 ± 0.0080
s4GAN + Entropy (Ours)	**0.398**	0.456	0.494
s4GAN + Margin (Ours)	0.360	**0.480**	**0.546**

Table 2: mIoU Scores for the UC Merced Land Use Classification Dataset [59]

Labeled Ratio(R)	2%	5%	12.5%
s4GAN [31] (Baseline)	0.403 ± 0.035	0.486 ± 0.018	0.511 ± 0.007
s4GAN + Entropy (Ours)	0.469	**0.513**	**0.554**
s4GAN + Margin (Ours)	**0.511**	**0.513**	0.529

Table 3: mIoU Scores for the DeepGlobe Land Cover Classification Dataset [9]

three experiments with different random seeds for robust evaluation for our baseline method. We compare our approach of using active learning to select representative labeled examples with this baseline. We experiment with two pool-based query strategies, *entropy* and *margin* sampling and demonstrate qualitative and quantitative performance improvements on two datasets, DeepGlobe Land Cover Classification Dataset [9], and UC Merced Land Use Classification Dataset [46, 59] over the stated baseline. We evaluated our approach with labeled ratios of 2%, 5%, and 12.5%. The qualitative results in Figures 3 and 4 are shown for the best out of the two sampling strategies for each labeled ratio. Table 1 shows the number of labeled images in each dataset for different labeled ratios.

UC Merced Land Use Classification Dataset: Table 2 shows a quantitative comparison of our method with the baseline for the UC Merced Land Use Classification Dataset [46, 59]. We compare the performance of entropy and margin sample selection strategies with the baseline and show significant and consistent performance improvements. Both the active learning strategies out-perform the baseline by a significant margin. We report a maximum mIoU improvement of close to 15% with as little as 2% labeled data, a maximum improvement of about 18% over the baseline when training with 5% and 12.5% labeled data across the two active learning strategies.

Figure 3 shows how our proposed method qualitatively improves over the UC Merced Dataset baseline for different labeled ratios. Our method predicts a finer coastline with no false positives, even with only as few as 34 labeled images which are 2% of labeled data (Row 1 of Figure 3). Similarly, we demonstrate that even when using only 5% (85 images) of labeled data (Row 2 of Figure 3), our method predicts the green-looking river that is camouflaging into the background while the baseline method completely misses it (Row 2 of Figure 3). This shows the importance of having a representative pool of labeled data, especially in a low data regime, as is our case. With 12.5% (211 images) of labeled data (Row 3 of Figure 3), our method accurately predicts the complex shape of the airplane (Column d), as opposed to the baseline (Column c), which was confused between multiple unrelated classes.

DeepGlobe Land Use Classification Dataset: Table 3 shows a quantitative comparison of our method with the baseline for the DeepGlobe Land Cover Classification Dataset [9]. We report significant performance improvements over the baseline using both entropy and margin sampling strategies. We report a maximum mIoU improvement of close to 27% with as little as 2% labeled data, a maximum improvement of about 6% over the baseline when training with 5% labeled data, and an improvement of approximately 8% with 12.5% labeled data across the two active learning strategies.

Figure 4 shows some visualizations from the DeepGlobe Dataset where it is seen that our method results in fewer false positives than the baseline.

5. Conclusion

This work proposes a method to leverage active learning-based sampling techniques to improve performance on the downstream task of semi-supervised semantic segmentation for land cover classification in satellite images. We do this by intelligently selecting samples for which pixel-wise labels should be obtained using coarse image classification-based active-learning strategies. Our method helps the semi-supervised semantic segmentation network start with an optimal set of labeled examples to help it get the right amount of initial information to learn the suitable representation. We prototype this method for a GAN-based semi-supervised semantic segmentation network, where the labeled images were selected using pool-based active learning strategies. We demonstrate the efficacy of our method for two satellite image datasets, both quantitatively and qualitatively, and report sizable performance gains.

Acknowledgements

References

[1] Les E Atlas, David A Cohn, and Richard E Ladner. Training connectionist networks with queries and selective sampling. In *Advances in neural information processing systems*, pages 566–573. Citeseer, 1990. 2

[2] Maria-Florina Balcan, Andrei Broder, and Tong Zhang. Margin based active learning. In *International Conference on Computational Learning Theory*, pages 35–50. Springer, 2007. 2

[3] Bindita Chaudhuri, Begüm Demir, Subhasis Chaudhuri, and Lorenzo Bruzzone. Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):1144–1158, 2017. 6

[4] Hao Chen and Zhenwei Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10):1662, 2020. 1

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 6

[6] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision, 2021. 3

[7] Ying Chen, Xu Ouyang, Kaiyue Zhu, and Gady Agam. Mask-based data augmentation for semi-supervised semantic segmentation. *arXiv preprint arXiv:2101.10156*, 2021. 2

[8] Tivadar Danka and Peter Horvath. modAL: A modular active learning framework for Python. available on arXiv at https://arxiv.org/abs/1805.00979. 6

[9] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 172–181, 2018. 2, 5, 8

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6

[11] M Dharani and G Sreenivasulu. Land use and land cover change detection by using principal component analysis and morphological operations in remote sensing applications. *International Journal of Computers and Applications*, pages 1–10, 2019. 1

[12] Huihui Dong, Wenping Ma, Yue Wu, Jun Zhang, and Licheng Jiao. Self-supervised representation learning for remote sensing image change detection based on temporal prediction. *Remote Sensing*, 12(11):1868, 2020. 1

[13] Geoffrey French, Samuli Laine, Timo Aila, Michal Mackiewicz, and Graham Finlayson. Semi-supervised semantic segmentation needs strong, varied perturbations. In *British Machine Vision Conference*, number 31, 2020. 2, 3

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27:2672–2680, 2014. 5

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016. 6

[16] Steven CH Hoi, Rong Jin, Jianke Zhu, and Michael R Lyu. Batch mode active learning and its application to medical image classification. In *Proceedings of the 23rd international conference on Machine learning*, pages 417–424, 2006. 2

[17] Wei-Chih Hung, Yi-Hsuan Tsai, Yan-Ting Liou, Yen-Yu Lin, and Ming-Hsuan Yang. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*, 2018. 1, 2, 3, 7

[18] Gong Jianya, Sui Haigang, Ma Guorui, and Zhou Qiming. A review of multi-temporal remote sensing data change detection algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B7):757–762, 2008. 1

[19] Ajay J Joshi, Fatih Porikli, and Nikolaos P Papanikolopoulos. Scalable active learning for multiclass image classification. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2259–2273, 2012. 2

[20] Benjamin Kellenberger, Diego Marcos, Sylvain Lobry, and Devis Tuia. Half a percent of labels is enough: Efficient animal detection in uav imagery using deep cnns and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):9524–9533, 2019. 2

[21] Jongmok Kim, Jooyoung Jang, and Hyunwoo Park. Structured consistency loss for semi-supervised semantic segmentation. *arXiv preprint arXiv:2001.04647*, 2020. 3

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[23] Adriana Kovashka, Olga Russakovsky, Li Fei-Fei, and Kristen Grauman. Crowdsourcing in computer vision. *arXiv preprint arXiv:1611.02145*, 2016. 1

[24] David D Lewis and Jason Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994*, pages 148–156. Elsevier, 1994. 2, 4

[25] David D Lewis and William A Gale. A sequential algorithm for training text classifiers. In *SIGIR'94*, pages 3–12. Springer, 1994. 2, 4

[26] Yaping Lin, George Vosselman, Yanpeng Cao, and Michael Ying Yang. Active and incremental learning for semantic als point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:73–92, 2020. 2

[27] Xiaoping Liu, Jialv He, Yao Yao, Jinbao Zhang, Haolin Liang, Huan Wang, and Ye Hong. Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science*, 31(8):1675–1696, 2017. 1

[28] Radek Mackowiak, Philip Lenz, Omair Ghori, Ferran Diego, Oliver Lange, and Carsten Rother. Cereals-cost-effective region-based active learning for semantic segmentation. *arXiv preprint arXiv:1810.09726*, 2018. 2

[29] Dwarikanath Mahapatra, Behzad Bozorgtabar, Jean-Philippe Thiran, and Mauricio Reyes. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 580–588. Springer, 2018. 2

[30] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014. 1

[31] Sudhanshu Mittal, Maxim Tatarchenko, and Thomas Brox. Semi-supervised semantic segmentation with high-and low-level consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 1, 2, 3, 5, 6, 7, 8

[32] Viktor Olsson and Wilhelm Tranheden. Consistency regularization for semantic segmentation. 2020. 2

[33] Viktor Olsson, Wilhelm Tranheden, Juliano Pinto, and Lennart Svensson. Classmix: Segmentation-based data augmentation for semi-supervised learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1369–1378, 2020. 2, 3

[34] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12674–12684, 2020. 2, 3

[35] Sakrapee Paisitkriangkrai, Jamie Sherrah, Pranam Janney, and Anton Van Den Hengel. Semantic labeling of aerial and satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(7):2868–2881, 2016. 2

[36] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 7

[37] Hiranmayi Ranganathan, Hemanth Venkateswara, Shayok Chakraborty, and Sethuraman Panchanathan. Deep active learning for image classification. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3934–3938. IEEE, 2017. 2

[38] Andrés C Rodríguez, Stefano D'Aronco, Konrad Schindler, and Jan D Wegner. Mapping oil palm density at country scale: An active learning approach. *Remote Sensing of Environment*, 261:112479, 2021. 2

[39] Dan Roth and Kevin Small. Margin-based active learning for structured output spaces. In *European Conference on Machine Learning*, pages 413–424. Springer, 2006. 2

[40] Soumya Roy, Asim Unmesh, and Vinay P Namboodiri. Deep active learning for object detection. In *BMVC*, page 91, 2018. 2

[41] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *arXiv preprint arXiv:1606.03498*, 2016. 5

[42] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53:197–207, 2019. 2

[43] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009. 2, 3

[44] H Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee proceedings of 5th annual workshop on computational learning theory, 287–294. *New York, ACM Press*, 10:130385–130417, 1992. 2

[45] Wei Shao, Liang Sun, and Daoqiang Zhang. Deep active learning for nucleus classification in pathology images. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 199–202. IEEE, 2018. 2

[46] Zhenfeng Shao, Weixun Zhou, Xueqing Deng, Maoding Zhang, and Qimin Cheng. Multilabel remote sensing image retrieval based on fully convolutional network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:318–328, 2020. 1, 2, 6, 8

[47] Asim Smailagic, Pedro Costa, Hae Young Noh, Devesh Walawalkar, Kartik Khandelwal, Adrian Galdran, Mostafa Mirshekari, Jonathon Fagert, Susu Xu, Pei Zhang, et al. Medal: Accurate and robust deep active learning for medical image analysis. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 481–488. IEEE, 2018. 2

[48] Nasim Souly, Concetto Spampinato, and Mubarak Shah. Semi supervised semantic segmentation using generative adversarial network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5688–5696, 2017. 1, 2, 3

[49] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 6

[50] André Stumpf, Nicolas Lachiche, Jean-Philippe Malet, Norman Kerle, and Anne Puissant. Active learning in the spatial domain for remote sensing image classification. *IEEE transactions on geoscience and remote sensing*, 52(5):2492–2507, 2013. 2

[51] Shuting Sun, Lin Mu, Lizhe Wang, and Peng Liu. L-unet: An lstm network for remote sensing image change detection. *IEEE Geoscience and Remote Sensing Letters*, 2020. 1

[52] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems*, pages 1195–1204, 2017. 3

[53] Marian Tietz, Thomas J. Fan, Daniel Nouri, Benjamin Bossan, and skorch Developers. *skorch: A scikit-learn compatible neural network library that wraps PyTorch*, July 2017. 6

[54] Devis Tuia, Frédéric Ratle, Fabio Pacifici, Mikhail F Kanevski, and William J Emery. Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2218–2232, 2009. 2

[55] Devis Tuia, Michele Volpi, Loris Copa, Mikhail Kanevski, and Jordi Munoz-Mari. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):606–617, 2011. 2

[56] Shuai Xie, Zunlei Feng, Ying Chen, Songtao Sun, Chao Ma, and Mingli Song. Deal: Difficulty-aware active learning for semantic segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 2

[57] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015. 6

[58] Lin Yang, Yizhe Zhang, Jianxu Chen, Siyuan Zhang, and Danny Z Chen. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 399–407. Springer, 2017. 2

[59] Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279, 2010. 1, 2, 5, 6, 8

[60] Qiangqiang Yuan, Huanfeng Shen, Tongwen Li, Zhiwei Li, Shuwen Li, Yun Jiang, Hongzhang Xu, Weiwei Tan, Qianqian Yang, Jiwen Wang, et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241:111716, 2020. 1

[61] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6023–6032, 2019. 3

[62] Bing Zhang, Di Wu, Li Zhang, Quanjun Jiao, and Qingting Li. Application of hyperspectral remote sensing for environment monitoring in mining areas. *Environmental Earth Sciences*, 65(3):649–658, 2012. 1