2021 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-4115-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIP42928.2021.9506533

SEMI-SUPERVISED LEARNING FOR MARS IMAGERY CLASSIFICATION

Wenjing Wang *Lilang Lin* *Zejia Fan* *Jiaying Liu* *

Wangxuan Institute of Computer Technology, Peking University, Beijing, China

ABSTRACT

With the progress of Mars exploration, numerous Mars image data are collected and need to be analyzed. However, because of the imbalance and distortion in Mars data, the performance of existing classification models is unsatisfactory. In this paper, we design a new framework based on semi-supervised contrastive learning for Mars rover image classification. The redundancy of Mars data can disable the effectiveness of contrastive learning. To strip out problematic learning samples, we propose to ignore inner-class pairs on labeled data as well as neglect negative pairs on unlabeled data. Experimental results show that our learning strategies can improve the classification model by a large margin and outperform state-of-the-art methods.

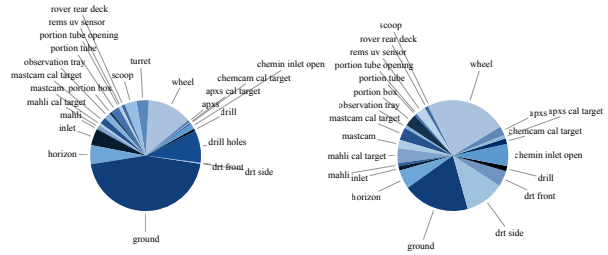
Index Terms— Mars image, classification, representation learning, semi-supervised learning, unsupervised learning

1. INTRODUCTION

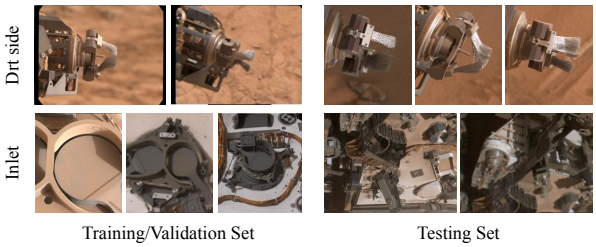
Extraterrestrial exploration is the forever frontier topic for human science research. In resent years, machine learning has shown its great power for planetary science, such as capturing the first black hole image [1]. With more space missions in the next ten years, machine learning will play a more important role in planetary science.

Mars Science Laboratory (MSL) rovers have been sending an enormous amount of data to earth. Organizing these massive images by content costs scientists a lot of time, which can be saved by autonomous classifiers. Convolutional Neural Networks (CNNs) have achieved great success in many image classification tasks [2, 3]. However, for Mars rover data, the performance of existing CNNs is unsatisfactory.

Generally, Mars rover data poses two challenges for classification: 1) **Train-test gap**. In extraterrestrial exploration, only past and current data can be used for training (and validation), while future data is the testing target. However, since MSL rovers do not collect images with equal frequency by every instrument on every sol, the data in training and testing varies. As shown in Figure 1, the gap between training and testing sets lies in not only the uneven distribution of categories, but also the various appearance of the target objects. 2) **Low image quality**: Because of the signal loss in Mars-to-Earth transmission or camera errors, the visual quality of some images may be damaged. As shown in Figure 2, images suffer from many kinds of distortion, including but not limited to low resolution, over-exposure, and color channel error.



(a) Category percentage difference



(b) Examples of object appearance difference

Fig. 1. The train-test difference on category distribution and object appearance in MSL rover data.

Fig. 2. Examples of low-quality images in MSL rover data.

For the problem of train-test gap, there has been a lot of research on data imbalance [4, 5, 6, 7] and classification generalization [8, 9, 10]. However, since the gap between training and testing on MSL rover data is too challenging and complex, these approaches are unsatisfactory.

Image visual quality usually can be improved by enhancement methods, such as SRCNN [11] for super-resolution.

* Corresponding author.

This work was supported by the National Key Research and Development Program of China under Grant No. 2018AAA0102702.

However, the image quality distortion on MSL rover data is a combination of multiple distortions, therefore too complex for existing enhancement methods to handle.

Different from existing methods, we propose to solve the problem by representation learning. By developing robust visual representation, the problems of train-test gap and low image quality can be solved all at once. Contrastive learning [12, 13] is a powerful self-supervised learning approach. In this paper, we extend it to help with the task of MSL rover image classification. The challenge in performing contrastive learning with MSL rover data is that, there is a severe information overlap between different MSL data samples, which can negate the effect of contrastive learning. To resolve this contradiction, we modify contrastive learning in two ways: *on labeled data*, we make use of annotations and propose to ignore inner-class pairs; *on unlabeled data*, we only maximize the similarity of positive pairs and do not consider the dissimilarity of negative pairs. Experimental results show that our semi-supervised classification framework outperforms existing methods by a large margin.

2. RELATED WORKS

2.1. Machine Learning for Planetary Science

Machine learning has been used for many planetary science tasks, such as exoplanet detection [14], comparative planetology and exoplanet biosignatures [15]. Readers may refer to [16] for a more comprehensive review and outlook.

For Mars exploration, Othrock *et al*. [17] and Wagstaff *et al*. [18] provide analysis for space and ground images. Wagstaff *et al*. [18] create a dataset of the Mars surface environment and train AlexNet [19] for content classification. However, the classification performance is unsatisfactory. In this paper, we have a deeper analysis of MSL rover data and design a more powerful classification framework.

2.2. Improving Classification Performance

Many techniques have been designed for improving classification performance. Several works [8, 9] propose loss designs to balance positive and negative samples. Classification problems often suffer from data imbalance across classes. To solve data imbalance, re-sampling [4] and re-weighting [6, 7] based methods are proposed. However, all these methods have limited effectiveness on MSL rover data. In this paper, we propose a more effective representation learning strategy and achieve superior performance.

3. SEMI-SUPERVISED MARS IMAGERY CLASSIFICATION

To solve the problems of MSL rover data, we extend contrastive learning into supervised inter-class and unsupervised similarity versions. Our framework is shown in Figure 3. In this section, we will first introduce the proposed learning strategies, then show our full framework architecture.

3.1. Supervised Inter-Class Contrastive Learning

Motivation. We first review the principle of contrastive learning. To learn feature representations, contrastive learning maximizes the agreement between different views of the same sample (positive pairs), and conversely minimizes the agreement between different samples (negative pairs). Given a sample x_i from the dataset X, contrastive learning first generates two views $d_1(x_i)$ and $d_2(x_i)$, where $d_1(\cdot)$ and $d_2(\cdot)$ are separate random data augmentation operators of the same augmentation family d. Denote F as the feature encoder and $z_i^j = F(d_j(x_i))$ as the extracted feature representation of $d_j(x_i)$, contrastive loss is:

$$\mathcal{L} = -\log \frac{\exp(\text{sim}(z_i^1, z_i^2)/\tau)}{\sum_{j \neq i} \exp(\text{sim}(z_i^1, z_j^2)/\tau)}, \tag{1}$$

where $\tau = 0.2$ is a temperature hyper-parameter, and $\text{sim}(u, v) = u^{\mathrm{T}} v / \|u\| \|v\|$ measures the similarity of two normalized vectors by dot production.

Contrastive learning is usually for self-supervised learning, *i.e*. training models from scratch by contrastive loss. But in this paper, we propose to use it in a semi-supervised way, *i.e*. training models with both supervised classification loss and contrastive loss. Taking advantage of the great power of contrastive learning, we enrich the visual representation and improve the robustness of the classification model, so that the problems of data imbalance and low image quality can be solved once and for all.

The essence of contrastive learning is maximizing the distance between different samples in the latent space. In ordinary large-scale classification datasets such as ImageNet [20], the difference between samples is usually big enough. However, Mars data is of less diversity: there is a significant information overlap between different samples. Rovers may take photos of the same scene multiple times, such as to check the status of equipment each day. Also, compared with the scenes on earth, Mars scenes are more monotonous. Because of the severe information overlap between different samples, contrastive learning turns out to be ineffective on mars data.

Combining Supervision. To resolve the contradiction between contrastive learning and MSL rover data, we introduce human supervision into contrastive learning. Using the classification annotations, we minimize the distance between samples only for samples belonging to the same class, and maximize the distance between samples only for samples belonging to different classes. In other words, we turn the originally *unsupervised sample-wise* contrastive learning into a *supervised inter-class* version. Denote x_{c_i} as a sample of class c_i,

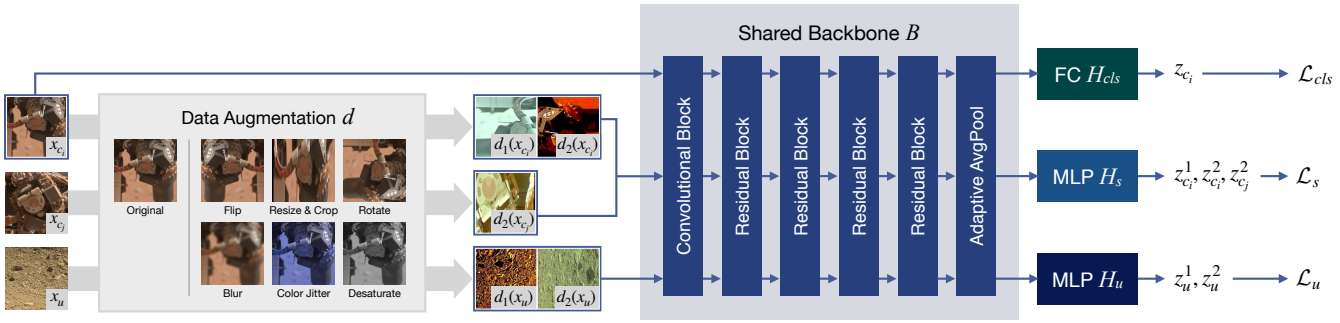


Fig. 3. The framework of our semi-supervised Mars image classification, which contains three streams: classification, supervised inter-class contrastive learning, and unsupervised similarity learning.

and x_{c_j} as a sample of class c_j, our supervised inter-class contrastive loss is:

$$\mathcal{L}_s = -\log \frac{\exp(\text{sim}(z^1_{c_i}, z^2_{c_i})/\tau)}{\sum_{c_j \neq c_i} \exp(\text{sim}(z^1_{c_i}, z^2_{c_j})/\tau)}. \quad (2)$$

Intuitively, the proposed $\mathcal{L}_s$ minimizes the distance between samples of the same class, and maximizes the distance between samples of different classes, which is similar to triplet loss [8] or center loss [10]. However, triplet loss and center loss do not enrich the feature representation, therefore have limited effectiveness as we will show in the experiments. In comparison, our supervised inter-class contrastive learning not only clusters the feature by categories, but also improves the feature representation, therefore better assists the classification task.

As shown in Figure 3, the data augmentation we use can be divided into two types: shape and pixel. Shape augmentation contains flip, random crop and resize, and rotation, which can help the model understand the object structure, and teach the model to recognize the same object under different angles and magnifications. Pixel augmentation contains Gaussian blur, color jittering, and desaturation, which can improve the model's robustness to image distortion and low visual quality.

3.2. Unsupervised Similarity Learning

Data annotation requires a large amount of human and financial resources. For Mars data, annotation is especially expensive. Recognizing the construction of the rovers, checking the status of the devices, and identifying Mars terrains all require expert knowledge. The lack of annotation limits the effectiveness of the proposed supervised inter-class contrastive loss. To further promote the performance, we propose to use cheap and easily available unlabeled data.

The aforementioned information overlap also exists in unlabeled data. What's worse, this time we have no human supervision to seek help from. To solve this dilemma, we propose to abandon negative pairs and only consider positive pairs. We call it similarity learning, since the model neglects the dissimilarity between samples. Denote x_u as a sample from unlabeled data, our unsupervised similarity loss is:

$$\mathcal{L}_u = -\text{sim}(z^1_u, z^2_u). \quad (3)$$

Discussion. Why $\mathcal{L}_u$ does not cause collapse is an interesting question. In experiments, we find that training with $\mathcal{L}_u$ alone does lead to collapse. But when we train together with $\mathcal{L}_s$, the collapse is prevented. This may be because our supervised inter-class contrastive learning restricts the feature representation to a good sub-space, which counteracts the bad impacts of unsupervised similarly learning.

3.3. Full Model

Our classification objective $\mathcal{L}_{cls}$ is the widely-used cross-entropy loss. Finally, our full objective is:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_s \mathcal{L}_s + \lambda_u \mathcal{L}_u, \quad (4)$$

where $\lambda_s = 1$ and $\lambda_u = 0.2$ control the balance of different training objectives.

In Figure 3, our backbone B is ResNet-50 [21]. For $\mathcal{L}_s$ and $\mathcal{L}_u$, we add 2-layer Multi-Layer Perceptron (MLP) heads H_s and H_u to project the representation into 128 dimensions.

4. EXPERIMENTS

4.1. Implementation Details

For model training and evaluation, we use the MSL Surface Dataset [18]. Wagstaff *et al*. [18] collect 6691 images by three instruments of the Curiosity rover. 24 categories are defined by a Mars rover mission scientist. Training and evaluation are split by day on Mars. Data on sol 3-564 is for training and validation, while sol 565–1060 is for testing. Different from [18], we reshuffle the training and validation sets. Our testing set remains the same as [18]. For unsupervised similarity learning, we additionally collect 34k unlabeled images from NASA's Planetary Data System (PDS).

Table 1. Results for MSL rover image classification.

Category	Method	Top-1 (%)
Baseline	AlexNet [19] in [18]	66.70
	ResNet-50 [21]	79.28 ± 1.76
Loss design	Triplet loss [8]	84.87 ± 1.13
	Center loss [10]	82.91 ± 0.93
	Focal loss [9]	82.86 ± 0.74
Semi-supervised learning	Pseudo labeling [22]	78.64 ± 0.04
	S4L [23], rotation	75.19 ± 1.73
	S4L [23], jigsaw	81.81 ± 2.33
Re-sampling	Decoupling [4], cRT	80.94 ± 1.43
	Decoupling [4], LWS	81.30 ± 0.62
Re-weighting	Class-balanced loss [6]	80.02 ± 0.89
	LDAM-DRW [7]	82.12 ± 1.92
Ours		**95.86** ± 1.63

The mini-batch size is 16 for $\mathcal{L}_{cls}$ and $\mathcal{L}_u$, and 24 for $\mathcal{L}_s$, which equals the number of categories. Our model is first pre-trained on ImageNet [20] by MoCo V2 [13], then fine-tuned on Mars data. The initial learning rate is 1e-3 for H_{cls} and 1e-6 for B, H_s and H_u. The optimizer is Adam [24]. We train for 30 epochs, with learning rate multiplied by 0.1 at 20 and 25 epochs, taking $\sim$1 hour with Nvidia GeForce RTX 2080Ti.

4.2. Comparison Results

Our model is compared with ten methods. For reliability, we run each experiment three times and report the mean and standard deviation. Results are shown in Table 1.

We first compare with the model in [18]. The reported performance is only 66.70%, which may be due to the limited capability of AlexNet [19]. When we change the baseline to ResNet-50 [21], the performance improves to 79.28%, demonstrating the necessity of using good feature extractors.

Next, we try to further promote the performance of ResNet-50 by three widely-used losses: Triplet loss [8], Center loss [10], and Focal loss [9]. Although these loss designs can improve the classification performance by 3$\sim$6%, their effectiveness is as not good as ours.

Since we introduce unlabeled data in our framework, we also test two semi-supervised learning methods. Pseudo learning [22] generates pseudo labels on unlabeled data. With pseudo learning, the performance degrades, which may be because the unlabeled data we crawled from PDS is uncurated. Compared with the MSL Surface Dataset, the collected unlabeled data is even more long-tailed and unbalanced. Therefore, the pseudo labels on unlabeled data can be wrong and mislead the classification model. S4L [23] applies self-supervised learning on unlabeled data. The jigsaw pretext task can improve the performance a little, while rotation declines the classification accuracy. This may be because most unlabeled data is about Martian soil and rock, which does not contain much semantic information. Therefore, unlabeled data is ambiguous for the jigsaw and rotation pretext tasks.

Table 2. Ablation studies of our designs.

Method	Top-1 (%)
ResNet-50 (baseline)	79.28 ± 1.76
MoCo V2 pre-training	81.56 ± 1.36
MoCo V2 pre-training + Strong Aug.	76.63 ± 3.06
MoCo V2 pre-training + Eq.(1)	75.66 ± 1.13
MoCo V2 pre-training + $\mathcal{L}_s$	93.82 ± 1.57
MoCo V2 pre-training + $\mathcal{L}_u$	78.65 ± 1.53
Final (MoCo V2 pre-training + $\mathcal{L}_s$ + $\mathcal{L}_u$)	**95.86** ± 1.63

We also compared with three state-of-the-art techniques for imbalanced data. Decoupling [4] is based on re-sampling. Class-balanced loss [6] and LDAM-DRW [7] are based on re-weighting. Although these techniques can improve the performance compared with the ResNet-50 baseline, their effectiveness is limited compared with our strategies.

Our Top-1 accuracy is 95.86%, higher than the second-best methods Triplet loss by a large margin of 10.99%. With our training strategies, the model learns better visual representation, thus has better generalization and robustness.

4.3. Ablation Studies

The demonstration of our designs is shown in Table 2. Thanks to the good feature learned by MoCo V2 [13] pre-training, the accuracy increases by 2.28%. The data augmentation we use in contrastive learning is stronger than normal data augmentations used for classification. If we directly apply our strong augmentation to classification, the performance degrades by 4.93%, which is because the supervision of classification is too weak to learn against strong augmentation.

Supervised inter-class contrastive learning $\mathcal{L}_s$ has a remarkable effect. If we use the original contrastive learning, *i.e*. Eq.(1), the accuracy declines, supporting our motivation of introducing inter-class supervision. Unsupervised similarity learning $\mathcal{L}_u$ alone degrades the performance, which is in line with our analysis in Section 3. Finally, our full model achieves the best performance, demonstrating the effectiveness of our semi-supervised learning framework.

5. CONCLUSION

In this paper, we propose a semi-supervised learning framework for MSL rover image classification. We extend contrastive learning into supervised inter-class and unsupervised similarity-only versions. Experimental results demonstrate the superiority of our designs.

6. REFERENCES

[1] Katherine L. Bouman, Michael D. Johnson, Daniel Zoran, Vincent L. Fish, Sheperd S. Doeleman, and William T. Freeman, "Computational imaging for VLBI image reconstruction," in *CVPR*, 2016, pp. 913–922.

[2] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.

[3] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015, pp. 1–9.

[4] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis, "Decoupling representation and classifier for long-tailed recognition," in *ICLR*, 2020.

[5] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen, "BBN: bilateral-branch network with cumulative learning for long-tailed visual recognition," in *CVPR*, 2020, pp. 9716–9725.

[6] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge J. Belongie, "Class-balanced loss based on effective number of samples," in *CVPR*, 2019, pp. 9268–9277.

[7] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Aréchiga, and Tengyu Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," in *NeurIPS*, 2019, pp. 1565–1576.

[8] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, 2015, pp. 815–823.

[9] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in *ICCV*, 2017, pp. 2999–3007.

[10] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao, "A discriminative feature learning approach for deep face recognition," in *ECCV*, 2016, vol. 9911, pp. 499–515.

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *TPAMI*, vol. 38, no. 2, pp. 295–307, 2016.

[12] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick, "Momentum contrast for unsupervised visual representation learning," in *CVPR*, 2020, pp. 9726–9735.

[13] Xinlei Chen, Haoqi Fan, Ross B. Girshick, and Kaiming He, "Improved baselines with momentum contrastive learning," *CoRR*, vol. abs/2003.04297, 2020.

[14] Christopher J. Shallue and Andrew Vanderburg, "Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90," *The Astronomical Journal*, vol. 155, no. 2, pp. 94, jan 2018.

[15] Sara I Walker, William Bains, Leroy Cronin, Shiladitya DasSarma, Sebastian Danielache, Shawn Domagal-Goldman, Betul Kacar, Nancy Y Kiang, Adrian Lenardic, Christopher T Reinhard, et al., "Exoplanet biosignatures: future directions," *Astrobiology*, vol. 18, no. 6, pp. 779–824, 2018.

[16] Abigail R. Azari, John B. Biersteker, Ryan M. Dewey, Gary Doran, Emily J. Forsberg, Camilla D. K. Harris, Hannah R. Kerner, Katherine A. Skinner, Andy W. Smith, Rashied Amini, Saverio Cambioni, Victoria Da Poian, Tadhg M. Garton, Michael D. Himes, Sarah Millholland, and Suranga Ruhunusiri, "Integrating machine learning for planetary science: Perspectives for the next decade," *CoRR*, vol. abs/2003.04297, 2020.

[17] Brandon Rothrock, Ryan Kennedy, Chris Cunningham, Jeremie Papon, Matthew Heverly, and Masahiro Ono, "Spoc: Deep learning-based terrain classification for mars rover missions," in *AIAA SPACE*, p. 5539. 2016.

[18] Kiri L. Wagstaff, You Lu, Alice Stanboli, Kevin Grimes, Thamme Gowda, and Jordan Padams, "Deep mars: CNN classification of mars imagery for the PDS imaging atlas," in *IAAI*, 2018, pp. 7867–7872.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NeurIPS*.

[20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li, "Imagenet: A large-scale hierarchical image database," in *CVPR*, 2009, pp. 248–255.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[22] Dong-Hyun Lee et al., "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *ICML Workshop*, 2013, vol. 3.

[23] Lucas Beyer, Xiaohua Zhai, Avital Oliver, and Alexander Kolesnikov, "S4L: self-supervised semi-supervised learning," in *ICCV*, 2019, pp. 1476–1485.

[24] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.