

LETTER

Unsupervised Building Damage Identification Using Post-Event Optical Imagery and Variational Autoencoder

Daming LIN[†], Jie WANG^{††}, *Nonmembers*, and Yundong LI^{†††a)}, *Member*

SUMMARY Rapid building damage identification plays a vital role in rescue operations when disasters strike, especially when rescue resources are limited. In the past years, supervised machine learning has made considerable progress in building damage identification. However, the usage of supervised machine learning remains challenging due to the following facts: 1) the massive samples from the current damage imagery are difficult to be labeled and thus cannot satisfy the training requirement of deep learning, and 2) the similarity between partially damaged and undamaged buildings is high, hindering accurate classification. Leveraging the abundant samples of auxiliary domains, domain adaptation aims to transfer a classifier trained by historical damage imagery to the current task. However, traditional domain adaptation approaches do not fully consider the category-specific information during feature adaptation, which might cause negative transfer. To address this issue, we propose a novel domain adaptation framework that individually aligns each category of the target domain to that of the source domain. Our method combines the variational autoencoder (VAE) and the Gaussian mixture model (GMM). First, the GMM is established to characterize the distribution of the source domain. Then, the VAE is constructed to extract the feature of the target domain. Finally, the Kullback–Leibler (KL) divergence is minimized to force the feature of the target domain to observe the GMM of the source domain. Two damage detection tasks using post-earthquake and post-hurricane imageries are utilized to verify the effectiveness of our method. Experiments show that the proposed method obtains improvements of 4.4% and 9.5%, respectively, compared with the conventional method.

key words: building damage identification, feature alignment, deep learning, domain adaptation

1. Introduction

In the past years, disasters have caused huge life losses and property damages worldwide. Thus, timely rescue operations are vital to mitigating damages. Building damage assessment can provide supportive information to facilitate rescue tasks. Damage assessment is traditionally conducted through ground survey, which is labor-intensive. Recently, the automatic interpretation of remote sensing data using machine learning has attracted significant attention.

Supervised machine learning algorithms, such as SVM [1] and KNN [2], have been investigated for building damage detection. Recently, deep learning has been

dominating various computer vision tasks, such as remote sensing classification [3], defect detection [4], etc. Consequently, deep learning was introduced into the area of building damage assessment. Vetrivel et al. proposed a damage identification method using convolutional neural networks (CNNs) and oblique aerial images, resulting in a classification accuracy of 91% [5]. Notably, many labeled samples are needed to train deep networks. However, generating sufficient samples from post-event images is difficult [6]. Li et al. proposed a semi-supervised method based on CNNs for post-hurricane damage detection, in which pre-training was used to decrease the requirement of labeled samples [6]. Some unsupervised learning techniques were also investigated. Li et al. proposed an unsupervised method to detect roof holes of rural area buildings from UAV images [7]. Moya et al. proposed a logistic regression based unsupervised learning model to identify collapsed buildings using satellite imagery [8].

As a branch of transfer learning, domain adaptation (DA) aims to transfer a classifier trained by samples of the source domain to the classification task of the target domain. Most DA methods pursue domain-invariant via minimizing the maximum mean discrepancy (MMD) or adversarial training using adversarial generative networks (GANs). MMD-based DA methods utilize MMD loss to reduce the distribution discrepancy between the features of the source and target domains. In the calculation of MMD loss, both source and target features are considered as holistic representations of the domains, while category-specific features are not fully considered. GAN-based methods use adversarial training to align the features of the source and target domains. The input feature is judged by the discriminator as either a source or target feature during adversarial training, while category-specific information is ignored. The lack of category-specific information during adaptation might cause mode collapse [9]. Mode collapse confuses the classification boundary, which leads to negative transfer. This issue should be given more attention when DA is applied to building damage detection because the image similarity of some categories is high, as illustrated in Figs. 2 and 3.

It is difficult to collect sufficient samples in the aftermath of disasters, which hinders the application of deep learning technology. Although DA provides a solution to reuse samples from different disaster scenes, it remains challenging due to that the conventional DA methods do not fully consider the category-level feature adaptation. Our motivation is to develop a novel DA framework which can

Manuscript received April 2, 2021.

Manuscript revised May 27, 2021.

Manuscript publicized July 20, 2021.

[†]The author is with Research Institute of Highway, Ministry of Transport, Beijing, 100088, Beijing, China.

^{††}The author is with School of Civil Engineering, North China University of Technology, Beijing, 100144, China.

^{†††}The author is with School of Information Science and Technology, North China University of Technology, Beijing, 100144, China.

a) E-mail: liyundong@ncut.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2021EDL8034

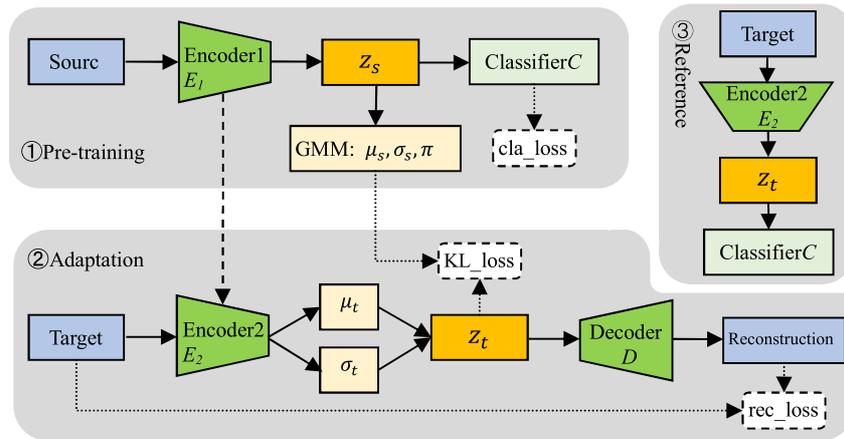


Fig. 1 Overview of our DA method which consists of three stages: pre-training, adaptation and reference. For conciseness, predictions of mean and standard variation are skipped in the reference stage.

conduct category-level feature alignment for complex image classification, such as building damage assessment. Inspired by Kingma’s work [10], we use the variational autoencoder (VAE) to estimate the distribution of the target domain. Meanwhile, the feature of each category is aligned with the corresponding category of the source domain via minimizing the KL divergence between the mixed Gaussian distributions of the two domains.

This research has two contributions. First, we propose a novel DA framework that combines the VAE and the GMM. Our DA algorithm can implement category-level alignment between the source and target domains. Second, we construct a building damage detection method based on the proposed algorithm using pose-event optical imageries.

The remainder of this letter is organized as follows. Section 2 describes our algorithm. Section 3 presents the experimental results. Section 4 discusses the method. Section 5 concludes this study.

2. Methodology

Our DA algorithm is shown in Fig. 1. The algorithm consists of three stages: pre-training, adaptation, and reference.

2.1 Pre-Training

The source domain networks consist of an encoder and a classifier, which are trained in a supervised way using labeled samples. The inputs are 200×200 images. Encoder 1 utilizes Resnet50 networks as the basebone. z_s refers to the feature of the source domain produced by Encoder 1. The dimension of the feature vector is 2048. The feature vectors are then fed into a classifier in which softmax is used to predict the category of each sample. Encoder 1 and Classifier C are trained together using cross-entropy error, defined as Eq. (1).

$$L_{src} = cls_loss = E_{(x_s, y_s) \sim (X_s, Y_s)} [C(E_1(x_s), y_s)] \quad (1)$$

where C is the source domain classifier, E_1 is Encoder 1,

x_s is the source domain sample, and y_s is the label. After training Encoder 1 and Classifier C, the GMM is estimated based on the feature extracted by Encoder 1. GMM includes three parameters: mean μ_s , standard variation σ_s , and π . $\pi = p(c)$ is the categorical distribution representing the probability of each category c . The number of Gaussian components equals the number of categories.

2.2 Adaptation

VAE is a generative model, which can generate samples from a specific probability distribution. In this work, we use the VAE to extract the feature of the target domain and force the feature’s distribution to approximate that of the source domain. From the perspective of generative model, the distribution of the source domain is used as a prior probability of the target domain.

The VAE of Fig. 1 includes Encoder 2 and a decoder. Encoder 2 has the same structure as Encoder 1. Encoder 2 is initialized by the weights of Encoder 1. Similarly, we use mixed Gaussian distributions to describe the feature space of the target domain. The Gaussian distribution parameters, namely, mean μ_t and standard variation σ_t , are predicted by Encoder 2, which is shown in Fig. 1. The feature of the target domain is conditioned on mean and standard variation using reparameterization, which is defined as Eq. (2).

$$z_t = \mu_t + \sigma_t \varepsilon \quad (2)$$

where $\varepsilon \sim N(0, 1)$ is a standard Gaussian distribution.

We use KL divergence to measure the discrepancy between the distributions of the source and target domains, which is defined as Eq. (3).

$$KL_loss = D_{KL}(q(z_t, c|x_t) || p(z_s, c)) \quad (3)$$

c denotes the category of the mixed Gaussian distributions. $p(z_s, c)$ is the joint probability distribution of the source domain. $q(z_t, c|x_t)$ can be factorized into $q(z_t|x_t)$ and $q(c|x_t)$ according to mean-field theory. Then, KL_loss is further

factorized as Eq. (4).

$$KL_{loss} = -E_{q(z_s, c|x_s)}[\log p(z_s|c) + \log p(c) - \log q(z_s|x_s) - \log q(c|x_s)] \quad (4)$$

In Eq. (4), $p(z_s|c)$ and $p(c)$ can be obtained from the GMM of the source domain, $q(z_s|x_s)$ is predicted by Encoder 2. $q(c|x_s)$ is the probability of the target sample belongs to category c . $q(c|x_s)$ is unknown because the samples of the target domain have no labels. To solve this issue, we adopt a mechanism called pseudo label in which the softmax output of Classifier C is used as $q(c|x_s)$. Thus, the KL_{loss} can be calculated. Refer to [11] for details.

In addition to the KL loss, the mean square error is used to reconstruct the target samples. rec_{loss} is defined as Eq. (5).

$$rec_{loss} = \frac{1}{n} \sum_{i=1}^n \|x_i - D(E_2(x_i))\|^2 \quad (5)$$

where E_2 stands for Encoder 2, and D is the decoder. Finally, the optimization objective of the adaptation process is defined as Eq. (6).

$$L_{tgt} = rec_{loss} + KL_{loss} \quad (6)$$

After training the VAE, Encoder E_2 and Classifier C are used for the damage detection task in the reference stage. Notably, Classifier C is reused because it remains unchanged during the adaptation stage, decreasing the training complexity.

3. Experimental Results

In the experiments, the samples of the target domain have no labels. Consequently, the damage detection task is carried out in an unsupervised manner. Our method is compared with some typical DA approaches. **Baseline**: a CNN classifier trained with only the data of the source domain. **DAN** [12]: DAN is a MMD-based method in which multiple task-specific layers are used to align features. **ADRF** [13]: ADRF is also a MMD-based domain adaptation method, which was proposed for the building damage detection of post-hurricane images. **ADDA** [14]: ADDA is a GAN-based method which can align features. **NO_KL**: KL loss is removed from Eq. (6), the autoencoder is trained only using the reconstruction error.

We implement the code with Pytorch 0.4.1. The GMM estimation is implemented with the Python Sklearn toolkit. The computer is equipped with Nvidia GTX1080TI to accelerate computation.

3.1 Post-Earthquake Datasets

In this study, the remote sensing images of the Haiti disaster make up the source domain, while the target domain includes images of Yushu earthquake. The images of the Haiti disaster can be found at www.haiti-patrimoine.org.



Fig. 2 Examples of datasets Haiti and Yushu. The first row denotes samples of Haiti dataset, the second row denotes Yushu dataset. The first and second columns indicate samples of undamaged buildings, the third and fourth are samples of partially damaged buildings.

Table 1 Comparison of classification results.

Methods	Haiti->Yushu		Sandy->Irma	
	mF1	Accuracy	mF1	Accuracy
Baseline	66.5	69.5	81.1	81.6
DAN	69.5	70.6	76.5	77.4
ADRF	70.5	71.9	81.3	84.1
ADDA	73.1	73.8	84.6	84.8
NO_KL	48.8	51.2	63.2	64.1
Ours	76.8	78.2	94.2	94.3

Images were segmented into many cells using the super-pixel method before training. Each cell was resized to a patch of 200×200 . After pre-processing, 4800 samples of the Haiti earthquake and 4800 samples of the Yushu earthquake were obtained. Some examples are shown in Fig. 2. The similarity between the partially damaged and undamaged buildings is high.

In this experiment, the samples are classified into three categories: undamaged, damaged, and others. The learning rates are set to $1e-4$ and $1e-5$ in the pre-training and adaptation stages, respectively. The SGD algorithm is used to optimize the networks. The classification results are shown in Table 1. Clearly, our method obtains the highest scores. Compared with ADDA, our method obtains accuracy improvements of 4.4%.

3.2 Post-Hurricane Datasets

The proposed algorithm is further validated using the post-hurricane datasets of the 2012 Sandy and 2017 Irma hurricanes. Dataset Sandy is used as the source domain, while dataset Irma is used as the target domain. Both domains have 5000 samples with a size of 200×200 . Some of the samples are shown in Fig. 3, which suggests that the similarity between categories is high.

The learning rate is set to $1e-4$ in the pre-training stage and $1e-6$ in the adaptation stage. The SGD optimizer with a momentum of 0.9 is used to update the network weights. The classification results are shown in Table 1 and indicate that our algorithm significantly outperforms the benchmarks. An improvement of 9.5% is obtained compared with the ADDA method. The experimental results without KL



Fig. 3 Examples of datasets Sandy and Irma. The first row denotes samples of Sandy dataset, the second row denotes images of Irma dataset. The first and second columns indicate samples of undamaged buildings, the third and fourth are the samples of damaged buildings.

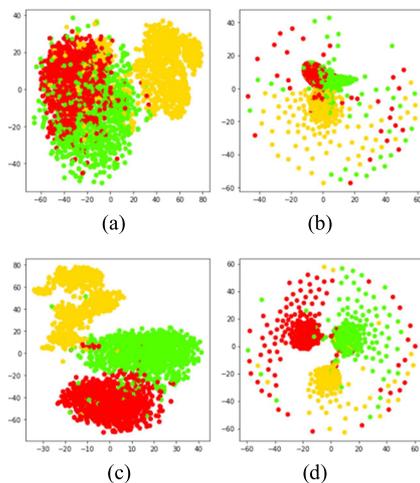


Fig. 4 Visualization of feature space. (a) and (c) are features of baseline for Yushu and Irma datasets, respectively. (b) and (d) are features of our method for Yushu and Irma datasets.

loss are also shown in Table 1. The performance deteriorates significantly without the KL constraint, proving that KL divergence minimization can reduce the cross-domain discrepancy.

4. Discussions

Separating damaged buildings from undamaged buildings is challenging due to the high similarity in image characteristics. To this end, we individually align each category via minimizing the KL loss to increase the inter-class distance in the feature space. Figure 4 visualizes the feature distributions of the target domain before and after adaptation. The features of most of the samples are more compact than those of the baselines. This effect can be attributed to the training strategy in which each category of the target domain is aligned with a specific Gaussian distribution in the source domain. In addition, the inter-class distances between categories are enlarged after adaptation, particularly for the task of Sandy->Irma. Notably, some outliers can be observed after adaptation, suggesting that these hard samples are different from those of the source domain.

5. Conclusion

Building damage assessment can help search teams during rescue operations when disasters strike. In this study, we present a novel DA framework to address the issue of insufficient category-specific information during feature adaptation. Our work is inspired by the VAE, which is used for reducing the cross-domain discrepancy via minimizing the KL divergence between two domains.

Two damage detection tasks using post-earthquake and post-hurricane datasets were investigated to validate the effectiveness of our method. The experimental results reveal that our method outperforms the state-of-the-art approaches. We also found that the outliers shown in Fig. 4 are difficult to be classified. In the future, we will investigate how to handle these difficult samples.

Acknowledgments

Dr. Daming Lin's work was supported by Science and Technology Research Project of Xinjiang Production and Construction Corps (Grant #2108AB028). Dr. Yundong Li's work was supported by National Natural Science Foundation of China (Grant #62071006).

References

- [1] J. Bialas, T. Oommen, U. Rebbapragada, and E. Levin, "Object-based classification of earthquake damage from high-resolution optical imagery using machine learning," *Journal of Applied Remote Sensing*, vol.10, no.3, p.036025, 2016.
- [2] M. Kakooei and Y. Baleghi, "Fusion of satellite, aircraft, and UAV data for automatic disaster damage assessment," *International Journal of Remote Sensing*, vol.38, no.8-10, pp.2511–2534, 2017.
- [3] X. Yu, X. Wu, C. Luo, and P. Ren, "Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework," *GIScience & Remote Sensing*, vol.54, no.5, pp.741–758, 2017.
- [4] Y. Li, W. Zhao, and P. Jiahao, "Deformable Patterned Fabric Defect Detection with Fisher Criterion based Deep Learning," *IEEE Transactions on Automation Science and Engineering*, vol.14, no.2, pp.1256–1264, 2017.
- [5] A. Vetrivel, M. Gerke, N. Kerle, F. Nex, and G. Vosselman, "Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.140, pp.45–59, 2018.
- [6] Y. Li, S. Ye, and I. Bartoli, "Semisupervised Classification of Hurricane Damage From Postevent Aerial Imagery Using Deep Learning," *J. Appl. Remote Sens.*, vol.12, no.4, p.045008, 2018.
- [7] S. Li, H. Tang, S. He, Y. Shu, T. Mao, J. Li, and Z. Xu, "Unsupervised Detection of Earthquake-Triggered Roof-Holes From UAV Images Using Joint Color and Shape Features," *IEEE Geoscience and Remote Sensing Letters*, vol.12, no.9, pp.1823–1827, 2015.
- [8] L. Moya, L.R.M. Perez, E. Mas, B. Adriano, S. Koshimura, and F. Yamazaki, "Novel Unsupervised Classification of Collapsed Buildings Using Satellite Imagery, Hazard Scenarios and Fragility Functions," *Remote Sensing*, vol.10, no.2, p.296, 2018.
- [9] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum Classifier Discrepancy for Unsupervised Domain Adaptation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2018)*,

- Salt Lake City, USA, 2018.
- [10] D.P. Kingma and M. Welling, "Auto-encoding variational Bayes," arXiv:1312.6114v10, May 2014.
- [11] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, "Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering," *IJCAI*, pp.1965–1972, 2017.
- [12] M.S. Long, Y. Cao, J.M. Wang, and M.I. Jordan, "Learning transferable features with deep adaptation networks," *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pp.97–105, 2015.
- [13] Y. Li, W. Hu, H. Li, H. Dong, B. Zhang, and Q. Tian, "Aligning Discriminative and Representative Features: An Unsupervised Domain Adaptation Method for Building Damage Assessment," *IEEE Transactions on Image Processing*, vol.29, pp.6110–6122, 2020.
- [14] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
-