

Ancient Architectural Damage Recognition Based on AlexNet and Transfer Learning

Liang Wang¹, Ke Ding²

¹College of Civil Engineering and Mechanics, Central South University of Forestry and Technology, Changsha 410004, China

²Hunan Provincial Key Laboratory of Engineering Rheology, Central South University of Forestry and Technology, Changsha 410004, China

Abstract: Due to the need for a large dataset to train image recognition using convolutional neural networks, obtaining accurately categorized images of ancient architectural structure damage poses certain challenges. To improve the recognition accuracy of damage in ancient architectural structures, we have enhanced the Vgg-16 network by combining transfer learning techniques with convolutional neural networks. This design, a new image recognition method based on transfer learning, effectively achieves precise identification of structural damage in ancient architecture, showing potential for real-world applications.

Keywords: Ancient architecture; Damage identification; Convolutional neural network; Transfer learning.

1. Introduction

Currently, a plethora of ancient architectures can be found across China. These edifices are not merely conglomerations of stones and earth. Instead, they stand as testimonies to history, bearers of culture, and emblems of national pride. However, as time progresses, many of these ancient buildings face grave threats of extensive damage and potential erasure. Their preservation is imperative, for if lost, the interwoven tapestries of culture and history they represent would be irretrievably erased as well. Ancient structures, already grappling with the wear and tear of age, continue to be eroded by natural disasters due to the influence of multifaceted environmental factors, invariably resulting in varying degrees of damage. To safeguard these invaluable structures for posterity, recognizing the damages in these ancient architectures is of paramount importance. There's a pressing need to periodically inspect and assess these structures, pinpointing the location and degree of damage, thereby preempting any catastrophic structural failures. Traditional methods for inspecting structural damages in ancient buildings chiefly encompass nondestructive testing techniques[1] and damage identification approaches predicated on structural dynamic characteristics.

With the advancement of artificial intelligence (AI) technology, numerous sectors have adeptly integrated AI to their advantage, and the realm of structural damage detection is no exception. Incorporating deep learning techniques into this domain has yielded commendable outcomes.[2] These deep learning techniques embrace, but are not restricted to, neural networks, support vector machines, random forests for damage detection, and data processing methods such as principal component analysis and independent component analysis. Particularly, neural networks—a vast interdisciplinary domain—have matured considerably in their application to detect damages in ancient structures and have proven especially conducive to inspecting these buildings. Conventional image processing algorithms fall short in delivering the required recognition efficacy for ancient architectural inspections. The advent of convolutional neural networks (CNNs) has substantially augmented the precision

of image recognition[3]. Convolutional neural networks, when trained on copious amounts of data, theoretically display enhanced learning outcomes and superior classification performance. Given the challenge in procuring datasets of damages in ancient architectures, scholars, striving to meet the precision demands for varying application scenarios, have introduced methods rooted in transfer learning to handle small-sample datasets. Transfer learning, a subdomain of machine learning, pivots on the central philosophy of applying knowledge acquired from one task to a new, albeit related, task, which in turn expedites and refines the learning process for the latter. Through this methodology, one can forgo starting model training from scratch. Especially when the data for the new task is scant, transfer learning can dramatically boost the learning efficiency and the resultant model's prowess.

2. Convolutional Neural Network (CNN)

2.1. Basic Principles of Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a deep learning architecture that has achieved tremendous success in the fields of computer vision and image processing. A typical CNN is composed of an input layer, hidden layers, and an output layer. Within the hidden layers, there are several components, including the convolutional layer (Conv), activation functions, pooling layer (Pool), and the fully connected layer (Full).

In the convolutional layer, the convolutional kernels (or filters) are responsible for capturing the local features of an image. Interestingly, the weights of these kernels are shared across different regions of the image. A single convolutional kernel is adept at extracting a specific feature from the input image. In contrast, multiple distinct kernels can collectively capture a wide range of features present in an image.

The computational formula for the convolutional layer is:

$$O(i,j)=\sum_{m=0}^{F-1}\sum_{n=0}^{F-1}I(i\cdot S+m,j\cdot S+n)\cdot K(m,n)+b$$

Among them: $O(i,j)$ is the element of the output feature map, $I(i \cdot S+m, j \cdot S+n)$ is the element of the input feature map, $K(m,n)$ is the weight of the convolutional kernel, and b is the bias term.

Since both the convolutional layer and the fully connected layer are linear operations, without an activation function, no matter how many layers the neural network has, it can only represent linear transformations. This significantly limits the expressive capability of the network. By introducing non-linear activation functions, the network can learn and represent more complex non-linear mappings. The activation layers in a Convolutional Neural Network (CNN) play a crucial role in this aspect. They typically involve a non-linear activation function, such as ReLU, tanh, or sigmoid, among others.

The pooling layer, also known as the subsampling layer, is introduced because the size of input images is typically large, and the feature maps obtained through convolution operations might also be extensive, leading to an increased computational complexity. The primary purpose of the pooling layer is to reduce the size of the feature maps, thus decreasing the computational demands and the number of parameters in the network. Fewer parameters mean that less

training data is required to train the network, simplifying the training process. This helps prevent overfitting, enhancing the model's generalization capability and increasing its robustness. In pooling operations, commonly used methods include max pooling and average pooling. If the max pooling function is utilized, the computation formula for the pooling layer is:

$$p_{pool} = \text{Max}(s_{ij}, s_{i+1,j}, s_{i,j+1}, s_{i+1,j+1})$$

Among them: p_{pool} is the result of pooling, s_{ij} is a positional element on the tensor of the feature map.

The fully connected layer acts as a bridge within the convolutional neural network. It receives the high-dimensional features extracted from the convolutional and pooling layers and further processes them through dimensionality reduction and non-linear transformations. This transformation enhances the network's expressive power, allowing it to capture complex patterns more effectively. After processing through the fully connected layer, the network can make precise categorical judgments on the input data, thus accomplishing classification tasks.

2.2. Building Convolutional Neural Network Models Based on Transfer Learning.

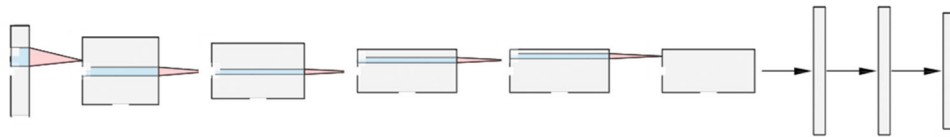


Figure 1. Structure of AlexNet.

AlexNet is a deep convolutional neural network (CNN) architecture that was proposed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton in 2012. It became renowned after winning the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) by a significant margin. This victory played a pivotal role in sparking interest in deep learning and CNNs within the broader machine learning and computer vision communities.

As the field of machine learning progresses, there's growing interest in how knowledge acquired in one domain can be transferred to another relevant domain. This has led to the study of Domain Adaptation, with the goal of adapting models between different domains. For instance, adapting a model trained for digit recognition to recognize handwritten fonts. With the emergence of deep learning, transfer learning has increasingly become a focal point of research. The advent of deep neural networks has made model transfer more flexible and effective. Transfer learning is not limited to domain adaptation; it encompasses various forms of knowledge transfer, such as feature extraction and parameter fine-tuning. Transfer learning has found extensive applications across various sectors, including computer vision, natural language processing, healthcare, and finance. It has achieved remarkable success in tasks like image classification, text classification, object detection, and speech recognition. While AlexNet was rigorously trained on the ImageNet dataset, which comprises over a million images, there isn't a sufficiently large dataset for images of ancient architectural damage. Hence, training a network model to achieve optimal classification performance for such images is challenging. Therefore, transfer learning emerges as an effective solution

for classifying datasets of ancient architectural structure damage.

3. Image Preprocessing

The experimental data come from the following two methods: ① Images sourced from online searches; ② Examples of ancient architectural structural damage photographed and collected using a mobile phone.

Due to the significant difference in image resolution between photos taken with mobile phones and those collected online, normalization is required. The data images are resized to $224 \times 224 \times 3$. The image data is split into training and test sets at a ratio of 7:3. Using a small sample dataset can lead to issues like overfitting, poor generalization capability, and difficulties in convergence. Thus, it's necessary to expand the dataset. Data augmentation operations are performed on the training images, including rotations of 90° , 180° , and 270° , mirroring, and translational transformations, ensuring that the original information of the image is retained to the maximum extent during this process.

4. Model Training and Prediction

This paper employs the AlexNet as its foundational architecture. AlexNet is a classic deep convolutional neural network pre-trained on the ImageNet dataset. We engage in transfer learning based on this model, fine-tuning the fully connected layers of the structure and introducing Dropout layers to prevent overfitting, catering to the task of identifying architectural damage. Simultaneously, the weights of the initial layers of the model are kept fixed, training only the

fully connected layers, which significantly accelerates the training speed while retaining the generic features learned from the ImageNet dataset. After preprocessing, the training dataset is employed to train the constructed transfer learning model. Optimizers continuously update and compute the network parameters to optimize the loss function to its extremum (either minimum or maximum). Throughout the training phase, our transfer learning model uses the Adam optimizer, which dynamically adjusts the learning rate for each parameter based on historical gradient information. An initial learning rate of 0.0001 is set and decayed as training progresses. During this phase, we log the loss and accuracy metrics of the model on both the training and validation datasets and plot these metrics over time, offering an intuitive insight into the model's training progression.

The study explores the performance and classification accuracy of the transfer learning model by observing the recognition accuracy on the validation set. Experiments are also conducted to investigate the effect of data augmentation on the recognition accuracy of the transfer learning model. Ultimately, the testing dataset evaluates the performance of the transfer learning model through key metrics such as training time and accuracy, verifying its precision in identifying architectural damage images. The results demonstrate that transfer learning in convolutional neural networks presents commendable recognition results for images of architectural structural damage.

5. Conclusion

In this study, we applied the image recognition techniques of convolutional neural networks to the task of identifying damage in ancient architectural structures using transfer

learning methods. The main findings of this paper are as follows:

① By performing essential preprocessing and enhancement on a small sample dataset and employing pre-trained models with transfer learning techniques, we specifically fine-tuned for the task of recognizing damage in ancient architectural structures. As the weights for most layers were frozen at the outset of training, only the fully connected layer's structure was adjusted. This approach significantly reduced the training time while ensuring the model's stability and accuracy.

② Through such training, the model successfully distinguished between damaged and intact images. The experimental results also demonstrated that its accuracy and efficiency in damage identification notably surpassed traditional image processing methods.

③ This paper presents an efficient solution for the practical application of convolutional neural networks combined with transfer learning methods in identifying damage in ancient architectural structures.

References

- [1] Li, G., et al., Stress wave velocity patterns in the longitudinal–radial plane of trees for defect diagnosis. *Computers and Electronics in Agriculture*, 2016. 124.
- [2] Li, C. and M. Wand, Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. *CoRR*, 2016. abs/1601.04589.
- [3] Wang, X., et al., Spectral prediction of *Phytophthora infestans* infection on tomatoes using artificial neural network (ANN). *International Journal of Remote Sensing*, 2008. 29(6): p. 1693-1706.