# EfficientNet-Based Transfer Learning for Modern Chinese Document Image Classification: A Case Study of Shanghai Library's "Modern Chinese Document Image Database" (Postprint)

**Authors:** Yang Min, Guo Limin

**Date:** 2023-09-05T00:00:00+00:00

## Abstract

Purpose/Significance: Images in modern historical documents, as important historical materials, have increasingly attracted the attention of humanities scholars. The deep annotation of large-scale image resources has consequently become a crucial component of image data infrastructure construction. Utilizing deep learning for content parsing of massive image collections represents a new direction in image research. The objective of this study is to address the automatic classification problem of large-scale modern document images and improve its accuracy and efficiency in practical applications through an empirical study of transfer learning based on EfficientNet for modern document image classification.

Methods/Process: The research methodology of this paper involves, based on feature analysis of modern document images, employing a dataset of 7,645 modern document images. Through serial superposition of image enhancement techniques such as cropping, white balance, color separation, and affine transformation, the diversity of sample images is increased. Furthermore, by investigating deep learning algorithms, transfer learning is performed using a fine-tuned simplified EfficientNet deep convolutional neural network model, ultimately obtaining a model that demonstrates good performance in modern document image classification.

Results/Conclusion: The research conclusion of this paper is that, based on experimental results, the model effectively improves image classification efficiency and accuracy, and possesses certain generalization value for addressing the automatic classification of large-scale images in modern historical documents.

# Full Text

## Preamble

**Title:** Transfer Learning Based on EfficientNet for Modern Document Image Classification: A Case Study of Shanghai Library's "Modern Chinese Literature Image Database"

**Authors:** YANG Min, GUO Limin (Shanghai Library, Shanghai 200031)

**Abstract:** *[Purpose/Significance]* Images in modern literature serve as crucial historical materials and are increasingly valued by humanities scholars. The deep annotation of large-scale image resources has become an essential component of image data infrastructure construction. Employing deep learning technologies to parse massive image collections represents a new direction in image studies. This research aims to address the challenge of automatic classification for large-scale modern document images through empirical investigation of transfer learning based on EfficientNet, thereby improving classification accuracy and efficiency in practical applications. *[Method/Process]* Our methodology involves analyzing characteristics of modern document images and employing a dataset of 7,645 modern literature images. We enhance sample diversity through serially 叠加 image augmentation techniques including cropping, white balance, tone separation, and affine transformations. By studying deep learning algorithms, we perform transfer learning using a fine-tuned simplified EfficientNet deep convolutional neural network model, ultimately obtaining a model that demonstrates strong performance in modern literature image classification. *[Results/Conclusions]* Experimental results reveal that this model effectively improves both classification efficiency and accuracy, offering significant potential for generalization in solving automatic classification challenges for large-scale modern document images.

**Keywords:** ResNet; EfficientNet; transfer learning; image classification; modern literature

**CLC Number:** TP391.41; G250.7

---

## 1 Introduction

Modern image resources have emerged abundantly in recent years, comprising historical photographs, illustrations from books and newspapers, and other documentary materials. In humanities research, images function as vital information carriers and have garnered increasing scholarly attention. As massive digital image resources continue to proliferate, their impact on knowledge production, dissemination, and humanities research grows ever more profound. Notable modern image digital repositories include the "National Newspaper Index: Modern Chinese Literature Image Database," the "Republican Era Image Resource Database" from National Library Publishing House, and the "Modern

Chinese Image Database" from China Social Sciences Press. Shanghai Library's "Modern Chinese Literature Image Database" stands out as particularly rich and valuable for research, currently containing 1 million images from modern periodicals and growing by 200,000 images annually.

With the advancement of convolutional neural network research, while accuracy has improved, network architectures have become increasingly large and dependent on massive datasets. Training a network from scratch incurs exceptionally high costs in architecture design and parameter tuning. Consequently, transfer learning has emerged as a popular approach for model construction. Deep learning-based image content parsing for large-scale resources represents a new frontier in image studies. Most existing cases focus on computer vision applications in biomedicine and other fields, with limited attention to historical image collections. Current image classification in these databases relies heavily on manual annotation, making it difficult to meet the demands of large-scale database construction and multi-dimensional user retrieval needs.

Recent years have witnessed remarkable breakthroughs in deep learning applications for face recognition, medical image analysis, and remote sensing image classification. Scholars have begun exploring these techniques for historical materials. For instance, Zheng Yuanpan et al. provided a comprehensive review of deep learning in image recognition, while Song Guanghui investigated image annotation methods based on transfer learning and deep convolutional features. In the library and information science domain, researchers have applied deep learning to address challenges in processing historical materials. Wang Haifeng et al. employed convolutional neural networks to cluster modern newspaper advertisement images, achieving nearly sixfold improvements in parsing efficiency. Yin Jieming et al. constructed a mobile visual search model for Dunhuang murals using deep learning and hashing methods. Gao Yaqi et al. proposed methods to improve image semantic classification accuracy through feature fusion. Yang Jianliang et al. noted that machine learning applications in archival management remain largely experimental, primarily employing supervised classification tasks on text and image data.

## 2 Related Research

Deep learning has become the predominant methodology in digital image classification research. By constructing deep neural network models, deep learning automatically learns features from digital images, eliminating the need for manual feature engineering that characterizes traditional approaches. Convolutional Neural Networks (CNNs) represent a crucial research direction within deep learning, demonstrating exceptional performance in image classification, processing, and audio-visual applications.

Transfer learning has become a hotspot in image classification applications. This approach leverages knowledge acquired from one domain to solve problems in another. In digital image classification, transfer learning typically employs pre-trained deep neural networks for feature extraction, then applies these features to new classification tasks. By using ImageNet-pretrained EfficientNet parameters as initial weights for our feature extraction model, we significantly reduce training time and computational resources while mitigating overfitting. Fine-tuning and data augmentation techniques further adapt the model to modern literature image classification tasks, substantially improving accuracy and robustness.

EfficientNet, introduced by Tan et al. in 2019, presents a multidimensional mixed model scaling method based on Neural Architecture Search (NAS). Unlike previous approaches that scaled single dimensions (depth, width, or resolution) arbitrarily, EfficientNet performs unified multidimensional scaling. While scaling from ResNet-18 to ResNet-152 by increasing network depth improves accuracy, such methods are difficult to optimize and cannot guarantee optimal performance. EfficientNet examines the interdependencies among network depth, width, and input resolution, discovering that these dimensions mutually influence one another. Through structured scaling that balances speed and accuracy, EfficientNet-B7 achieved 84.3% Top-1 and 97.1% Top-5 accuracy on ImageNet-1K, surpassing GPipe while using 8.4 times fewer parameters and operating 6.1 times faster. Compared to ResNet-50, EfficientNet-B4 improved Top-1 accuracy from 76.3% to 83.0% (+6.7%) with similar FLOPS.

However, the complex internal structure of MBConv modules makes EfficientNet computationally intensive and demanding on hardware resources, with lengthy training times. For smaller datasets like modern literature images, complex networks risk overfitting. Wu Suwen et al. applied transfer learning-based image classification to poetry imagery, collecting and organizing nine major categories of poetic imagery data and migrating ImageNet-trained EfficientNet models to this domain, achieving satisfactory classification results that met project requirements for poetry-based image search.

## 3 Theoretical Model and Methods: Convolutional Neural Networks and EfficientNet Selection

Convolutional Neural Networks represent one of the most commonly used neural network architectures in deep learning, primarily comprising convolutional layers, pooling layers, and fully connected layers.

**Convolutional Layers:** The core component of CNNs, convolutional layers learn multiple filters to extract diverse features. Each filter functions as a small neural network that slides across digital images to generate feature maps. Multiple filters learn different features, enabling rich feature extraction from digital images.

**Pooling Layers:** These layers reduce feature map dimensions and decrease neural network parameters. Common pooling methods include max pooling (selecting maximum values from each patch) and average pooling (computing mean values). This reduction helps control computational complexity.

**Fully Connected Layers:** The final layers of the network convert feature maps into classification results through one or more fully connected layers, typically using softmax functions to produce probability distributions across categories.

**EfficientNet Architecture:** EfficientNet's efficiency depends critically on its baseline network. The B0 baseline network, discovered through neural architecture search, contains 9 stages. Stage 1 consists of a 3$ $\times 3 convolutionlayerwithstride2, batchnormalization, andSwishactivation.Stages2-8comprise16MBConv(MobileInvertedBottleneckConvolution)modules, eachcontainingdepthwiseseparablec$ $and - Excitation)channelattentionmechanisms.Stage9includesa1\times$ $1 convolution layer, global average pooling, and classification layers.

[Figure 1: see original paper] summarizes the structured scaling approach and B0 baseline architecture, where $d$, $w$, and $r$ represent scaling coefficients for depth, width, and resolution respectively.

[Figure 2: see original paper] illustrates the EfficientNet family scaling, showing how B1-B7 extend the baseline B0 through compound scaling coefficients. This approach achieves state-of-the-art accuracy on five public datasets while reducing parameters up to 21-fold compared to existing ConvNets.

## 4 Research Framework Design and Experiments

Our dataset comprises selected images from Shanghai Library's "Modern Chinese Literature Image Database," which integrates curated images from modern literature. Referencing domestic and international image classification standards, the database establishes 16 primary categories including photographs, paintings, calligraphy, woodcuts, manuscripts, cartoons, maps, charts, and others, with metadata cataloging at the element level.

The database contains 1 million images from modern periodicals, growing by 200,000 images annually. However, these images exhibit characteristics that challenge conventional deep learning approaches: relatively low resolution, limited color information (predominantly black-and-white), and small dataset scale. Current image feature extraction research focuses on modern high-resolution color images, making direct application to modern journal images problematic for feature extraction accuracy.

For this study, we randomly sampled 7,645 images. Core metadata includes title, creator, source publication, image type, publication year, volume/issue, size, and dimensions (see Table 1 and Table 2).

**Data Augmentation:** Raw images contain interference factors such as journal names and dates. We cropped these elements during annotation. Images from multiple batches exhibit varying color casts due to different acquisition environments. We applied white balance algorithms to simulate human visual adaptation, adjusting color balance toward true tones. Posterization reduced color channel bit depth to help networks recognize images from contours. Affine transformations (horizontal/vertical shearing) created deformed variants without destroying original features. Through serial stacking of these techniques, we expanded our dataset from 7,645 to 15,290 images, all resized to 224$×$224 pixels.

**Transfer Learning:** We employed transfer learning using ImageNet-pretrained EfficientNet parameters as our initial model weights. ImageNet provides tens of millions of cleanly labeled, quality-controlled images organized by WordNet hierarchy. This approach saved substantial training time and computational resources while reducing overfitting.

**Model Simplification:** The original EfficientNet's 16 MBConv modules proved overly complex for our modest dataset, risking overfitting. Through experimentation, we simplified the network to 6 MBConv modules (four 3$×3*modulesandtwo*5×$5 modules), preserving the core architecture while reducing parameters and computation. This simplified network achieved excellent classification results for modern literature images.

**Experimental Setup:** We randomly divided the augmented dataset into training (70%, 10,703 images), validation (20%, 3,058 images), and test (10%, 1,529 images) sets, employing 10-fold cross-validation.

## 5 Experimental Results Analysis

Using the simplified EfficientNet model with transfer learning and data augmentation, our 10-fold cross-validated model achieved 90.97% average Top-1 accuracy and 91.00% average F1 score, demonstrating simplicity, efficiency, and strong generalization capability for modern literature image classification.

Table 3 presents detailed classification performance across 16 categories. The simplified EfficientNet achieved over 85% accuracy for all categories, with precision, recall, and F1 scores exceeding 84% except for woodcuts, rubbings, and manuscripts.

We employed confusion matrix analysis to visually examine inter-category misclassification (Table 4). Rows represent true categories while columns represent predicted categories. We observed that manuscripts and calligraphy exhibit high morphological similarity and phototropism, creating challenges even for human annotators. This similarity prevented the network from effectively learning discriminative features, resulting in suboptimal classification. However, this insight provides valuable guidance for future model optimization.

Comparative experiments with ResNet50_{vd} on the identical dataset and training environment demonstrated that our simplified EfficientNet outperformed ResNet50_{vd} across all metrics: 2.76% higher Top-1 accuracy, 1.57% higher precision, 1.98% higher recall, and 2.36% higher F1 score, with significantly improved computational efficiency. These results confirm that our approach can economically and efficiently support incremental model training for high-precision AI classification of modern literature image databases.

## 6 Conclusion and Future Work

This study, based on 7,645 image samples, demonstrates that transfer learning with a simplified EfficientNet model achieves 90.97% Top-1 accuracy and 91.00% F1 score through data augmentation. The model exhibits simplicity, efficiency, and robust generalization for modern literature image classification. Performance gains over ResNet50-vd confirm its superiority and computational advantages.

Future research will explore applications in digital image semantic information extraction. Through preprocessing and extraction of color, shape, texture, and contextual features, we aim to enable automatic extraction of most semantic information, truly reducing manual intervention and achieving semantic description of million-scale image data.

Modern periodical images primarily function as illustrations rather than independent historical materials, bearing strong relationships with accompanying text. Current metadata relies heavily on this text, with limited semantic conceptual depth. For individual images, semantics remain sparse without textual context. We envision combining EfficientNet-based transfer learning with text analysis to maximize semantic extraction from both visual and textual information, bridging the gap between low-level visual features and high-level semantics. This represents our primary research direction moving forward.

## References

[1] ZHENG Y P, LI G Y, LI Y. Survey of application of deep learning in image recognition [J]. Computer engineering and applications, 2019, 55(12): 20-36.

[2] SONG G H. Image annotation method based on transfer learning and deep convolutional feature[D]. Hangzhou: Zhejiang University, 2017.

[3] WANG H F, DENG K, CHEN J. Clustering methods of modern newspaper advertisements via convolutional neural network[J]. Digital humanities, 2021(2): 50-61.

[4] YANG M, XIA C J, YAN J. Analysis of the current situation and prospect of image database construction from the perspective of digital humanities[J]. Library journal, 2021, 40(4): 90-99.

[5] YIN J, ZENG Z M, SUN S Q. Research on the mobile visual search of Dunhuang murals based on deep learning and hashing[J]. Journal of modern information, 2023, 43(5): 35-45, 78.

[6] CHU M N. Research of image classification technology based on convolutional neural network[D]. Xiangtan: Xiangtan University, 2015.

[7] GAO Y Q, WANG H, LIU Y C. A study on the exploring of image semantic features and their influence on classification[J]. Information science, 2021, 39(10): 107-117.

[8] MA Y C, LIU Y J, XIE Q, et al. Review of automatic image annotation technology[J]. Journal of computer research and development, 2020, 57(11): 2348-2374.

[9] YANG J L, LIU Y N. The application of machine learning in archives management: Progress and challenges[J]. Archives science bulletin, 2019(6): 48-56.

[10] YANG Z Z, KUANG N, FAN L, et al. Review of image classification algorithms based on convolutional neural networks[J]. Journal of signal processing, 2018, 34(12): 1474-1489.

[11] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2014: 1717-1724.

[12] SCHUETTPELZ E, FRANDSEN P, DIKOW R, et al. Applications of deep convolutional neural networks to digitized natural history collections[J]. Biodiversity data journal, 2017, 5: e21139.

[13] WU S W, ZHAO H J, LIU X, et al. Research on application of image classification based on transfer learning in poetry[J]. Computer technology and development, 2021, 31(7): 215-220.

[14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2023-05-25]. http://arxiv.org/abs/1409.1556.

[15] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[EB/OL].[2023-04-09]. https://doi.org/10.48550/arXiv.1905.11946.

[16] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2009: 248-255.

[17] CHEN T, SHAN R R, LI H. Semantic annotation of image resources in digital humanities[J]. Journal of library and information science in agriculture, 2020, 32(9): 6-14.

[18] CHEN J J. A review of image semantic annotation research[J]. Research on library science, 2017(18): 2-7, 20.

---

**Table 1** Image sample structure

**Table 2** Core metadata and bibliographic information sources of image content

**Table 3** Classification performance of simplified EfficientNet network in 16 image classifications

**Table 4** Confusion matrix of simplified EfficientNet network in modern literature classification

*Note: Figure translations are in progress. See original paper for figures.*