Mining Security Assessment in an Underground Environment using a Novel Face Recognition Method with Improved Multiscale Neural Network

Authors: Xinhua Liu, Peng Qi, Patrick Siarry, Dezheng Hua, Z Ma, Xiaoqiang Guo, Orest Kochan, Z. Li, Z. Li

Date: 2024-04-01T00:00:00+00:00

Abstract

Excessive staffing in underground coal mining operations complicates daily management, while incomplete personnel information impedes rescue efforts during mine accidents. To address this critical safety issue, this paper proposes a novel face recognition method based on an improved multiscale neural network. Specifically, a new depthwise separable (DS)-Inception block is designed, and a joint supervised loss function based on center loss theory is developed to construct a new multiscale model. This enables recognition of miners in harsh underground environments during emergency rescue operations. Experimental results demonstrate that the proposed method achieves accuracy, recall, and F1-score metrics of 97.26%, 94.17%, and 95.42%, respectively, for miner face recognition in underground mining environments. The transfer model with joint supervised loss effectively improves recognition accuracy by approximately 0.5–1.5%. Furthermore, the proposed face recognition method attains an average recognition accuracy of 91.34% with a miss detection rate below 5% in coal mine excavation tunnels.

Full Text

Preamble

Mining Security Assessment in an Underground Environment using a Novel Face Recognition Method with Improved Multiscale Neural Network

Xinhua Liu[1], Peng Qi[1], Patrick Siarry[2], Dezheng Hua[1], Z. Ma[3], Xiaoqiang Guo[1], Orest Kochan[4], Z. Li[5]*

[1][Affiliation not provided]
[2][Affiliation not provided]
[3]Sustainable Buildings Research Centre, University of Wollongong, 2522, Australia
[4]Department of Telecommunications, Lviv Polytechnic National University, Bandery 12, 79013 Lviv, Ukraine
[5]Faculty of Mechanical Engineering, Opole University of Technology, 45-758 Opole, Poland

*Corresponding author

Abstract

Overstaffing in underground coal mining complicates daily management, and incomplete information about miners hinders rescue efforts during mine accidents. To address this critical safety and sustainability issue, this paper proposes a novel face recognition method based on an improved multiscale neural network. We design a new depthwise separable (DS)-Inception block and develop a joint supervised loss function grounded in center loss theory to construct a new multiscale model. This enables reliable miner recognition in harsh underground environments during rescue operations. Experimental results demonstrate that the proposed method achieves accuracy, recall, and F1-score of 97.26%, 94.17%, and 95.42%, respectively, for miner face recognition in underground mining environments. The transfer model with joint supervised loss effectively improves recognition accuracy by approximately 0.5–1.5%. Furthermore, the proposed face recognition method attains an average recognition accuracy of 91.34% with a miss detection rate below 5% in coal mine excavation tunnels.

Keywords: Mining Security; Coal Safety Assessment; Artificial Neural Network; Transfer Learning

1. Introduction

With the rapid advancement of coal mine informatization [1, 2], deep learning-based recognition technology [3, 4] has attracted considerable attention for underground mining safety assessment. Compared with traditional miner management systems [5, 6], underground coal mine face recognition systems [7] can provide timely, comprehensive, and reliable identification information to mine personnel management agencies while supplying rescuers with identity and location data for trapped miners during accidents. This technology plays a crucial role in curbing underground overproduction, strengthening mine management [8], and facilitating emergency rescue operations. However, the underground coal mine environment suffers from poor lighting conditions, and the mining process generates substantial dust, steam, and coal ash, which significantly complicate face recognition tasks. To address these challenges, this paper presents an optimized face recognition algorithm to improve system accuracy in underground coal mines.

Traditional approaches for occluded face recognition rely on sparse representation methods. Sparse Representation Classification (SRC) [9] represents high-dimensional images in a low-dimensional space, aiming to use the minimum number of training samples while achieving minimal fitting error. He et al. [10] proposed a sparse representation algorithm based on maximum entropy criterion that effectively handles non-Gaussian errors and outliers. To encode more structural and discriminative information, Zheng et al. [11] integrated adaptive learning weights into a Group Sparse Representation Classifier (GSRC). Chen et al. [12] introduced a nuclear norm-based matrix regression (NMR) method to mitigate the effects of contiguous occlusion on face recognition. Nevertheless, traditional face recognition methods have limited feature extraction capabilities, particularly when occlusion mixes with normal features, substantially reducing recognition effectiveness. Consequently, improving face recognition accuracy remains an urgent priority.

Deep learning advances [13, 14] have dramatically enhanced network feature extraction capabilities through model deepening. Wieczorek et al. [15] proposed a lightweight convolutional neural network for face detection in risk situations. Zhao et al. [16] developed a deep age-invariant model (AIM) for wild face recognition by combining age-invariant feature extraction and face feature synthesis. Qiu et al. [17] introduced a novel face recognition method using a single end-to-end deep neural network with strong anti-occlusion capability to identify and clean corrupted features. To further mitigate resolution discrepancies caused by imaging limitations, Gao et al. [18] proposed hierarchical deep CNN feature set-based representation learning for face recognition. However, current occluded face recognition research primarily focuses on specific facial region occlusions [19–21], such as eyes and mouth, and many models fail to achieve optimal performance across all face datasets.

Recent deep learning progress has substantially improved face recognition model adaptability. Transfer learning [22] addresses poor model generalization by applying knowledge from related tasks to neural network training. Cai et al. [23] proposed a generative adversarial network (OA-GAN) for natural face de-occlusion without requiring occlusion masks. Zhang et al. [24] developed a method to improve masked face recognition performance. Shukla et al. [25] employed MobileNet V2 with transfer learning for masked face identification and verification. Tang et al. [26] proposed a depth map transfer learning method for face recognition in unrestricted environments. However, public face datasets lack samples from coal mine environments, necessitating the creation of a miner face dataset for transfer model fine-tuning.

This paper proposes an improved face recognition method for underground coal dust occlusion based on transfer learning to address random coal dust obscuration. First, we design a novel DS-Inception block to reduce model parameters and establish a multiscale neural network named DSR-Inception. Second, we propose a joint supervised loss function combining center loss and softmax loss for face recognition classification tasks. Experimental results demonstrate that the proposed network outperforms classical face recognition models on our homemade miner face dataset in terms of accuracy, recall, and F1-score, and that transfer models with joint supervised loss achieve higher recognition accuracy. Finally, industrial tests in coal mine excavation tunnels validate the proposed algorithm’ s effectiveness.

The remainder of this paper is organized as follows: Section 2 introduces the proposed improved algorithm and model architecture. Section 3 presents comparative experiments on model transfer strategies and verifies the improved algorithm using the homemade miner face dataset. Section 4 describes the constructed face recognition system and occluded face recognition experiments in underground coal mines. Section 5 concludes the paper and outlines future work.

2.1 Transfer Learning

Transfer learning applies networks trained on related tasks to new tasks, addressing the limited generalization capability of traditional machine learning. In transfer learning, the domain $\mathcal{D}$ can be expressed by Equation (1):

$$\mathcal{D} = \{\mathcal{X}, P(X)\} \tag{1}$$

where $\mathcal{X}$ is the feature space, X represents sample data points with $X = \{x_1, x_2, \dots, x_n\}$, x_i is a feature vector, and $P(X)$ is the marginal probability.

The task $\mathcal{T}$ can be expressed by Equation (2):

$$\mathcal{T} = \{\mathcal{Y}, P(Y|X)\} \tag{2}$$

where $\mathcal{Y}$ is the feature space and $P(Y|X)$ is the objective function.

Based on these definitions, transfer learning utilizes knowledge from an existing source task $\mathcal{T}_s$ in a source domain $\mathcal{D}_s$ to solve a learning task $\mathcal{T}_t$ in a target domain $\mathcal{D}_t$, thereby achieving a better conditional probability distribution $P(Y_t|X_t)$ for the target domain.

2.2 Inception Block and Depthwise Separable Convolution

As shown in Appendix 1, the Inception block [27] is the fundamental convolutional building block of GoogleNet [28], which increases network width by splitting traditional convolutional kernels into multiple sizes. This enables more comprehensive feature extraction while consuming fewer computational resources. The feature maps F' obtained after the Inception block can be mathematically expressed by Equations (3)-(7):

$$F_1 = \mathrm{ReLU}(\mathrm{conv}(F, k_{1\times1}) + b_1) \tag{3}$$

$$F_2 = \mathrm{ReLU}(\mathrm{conv}(\mathrm{ReLU}(\mathrm{conv}(F, k_{1\times1})), k_{3\times3}) + b_2) \tag{4}$$

$$F_3 = \mathrm{ReLU}(\mathrm{conv}(\mathrm{ReLU}(\mathrm{conv}(F, k_{1\times1})), k_{5\times5}) + b_3) \tag{5}$$

$$F_4 = \mathrm{ReLU}(\mathrm{conv}(\mathrm{MaxPool}(F, k_{3\times3}), k_{1\times1}) + b_4) \tag{6}$$

$$F' = \mathrm{Concat}(F_1, F_2, F_3, F_4) \tag{7}$$

where F_1, F_2, F_3, and F_4 are feature maps from the four branches, $k_{i\times i}$ denotes convolution kernels of size $i \times i$, and b_i represents bias terms.

Depthwise separable convolution [29] combines two convolution types: depthwise convolution and pointwise convolution. While traditional convolution fuses channel and spatial information simultaneously, depthwise separable convolution isolates and processes these sequentially before fusion.

The operation occurs in two steps: depthwise and pointwise convolution. In the depthwise step, convolution operates on each image channel using single-layer planar kernels to obtain planar convolution results. In the pointwise step, these results are stitched together and processed using 1×1 pointwise convolution kernels.

The output feature map for conventional convolution (assuming stride one and padding) is expressed by Equation (8):

$$G_{k,l} = \sum_{i,j} K_{i,j} \cdot F_{k+i-1,l+j-1} \tag{8}$$

where G is the output feature map, F is the input feature map, and K is the conventional convolution kernel. The computational cost of conventional convolution is expressed by Equation (9):

$$\mathrm{Cost}_K = D_K \times D_K \times M \times N \times D_F \times D_F \tag{9}$$

where $D_K \times D_K$ is the kernel’ s spatial dimension, M is the number of input channels, N is the number of output channels, and $D_F \times D_F$ is the feature map size.

The output feature map of depthwise separable convolution with equivalent parameters is expressed by Equation (10):

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m} \tag{10}$$

where $\hat{G}$ is the output feature map, F is the input feature map, and $\hat{K}$ is the depthwise separable convolution kernel. The computational cost is expressed by Equation (11):

$$\text{Cost}_{\hat{K}} = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \tag{11}$$

which represents the sum of depthwise and 1×1 pointwise convolution costs. Replacing traditional convolution with depthwise separable convolution reduces network size and computational effort, as expressed by Equation (12):

$$\frac{\text{Cost}_{\hat{K}}}{\text{Cost}_K} = \frac{1}{N} + \frac{1}{D_K^2} \tag{12}$$

2.3 Improved DS-Inception Block

This paper fuses depthwise separable convolution into the Inception block by replacing traditional convolution kernels with depthwise separable ones, creating the improved DS-Inception block (structure shown in Appendix 2). The feature maps F' obtained after the DS-Inception block can be mathematically expressed by Equations (13)-(18):

$$F_1 = \text{ReLU}(\text{conv}(F, k_{1\times1}) + b_1) \tag{13}$$
$$F_2 = \text{ReLU}(\text{Dw}(\text{ReLU}(\text{conv}(F, k_{1\times1}))) + b_2) \tag{14}$$
$$F_3 = \text{ReLU}(\text{Dw}(\text{ReLU}(\text{conv}(F, k_{1\times1}))) + b_3) \tag{15}$$
$$F_4 = \text{ReLU}(\text{conv}(\text{MaxPool}(F, k_{3\times3}), k_{1\times1}) + b_4) \tag{16}$$
$$F_5 = \text{ReLU}(\text{conv}(F, k_{1\times1}) + b_5) \tag{17}$$
$$F' = \text{Add}(\text{Concat}(F_1, F_2, F_3, F_4), F_5) \tag{18}$$

where F_1, F_2, F_3, F_4, and F_5 are feature maps from the five branches, $k_{i\times i}$ denotes convolution kernels of size $i \times i$, and b_i represents bias terms.

Depthwise separable convolution replaces large kernels in the Inception block, and residual structure is incorporated. Research shows that residual structure [30] effectively mitigates gradient dispersion and accelerates model convergence. The improved block significantly reduces parameters compared to the original: with K channels per branch convolution kernel and N input channels, the original block has $4NK + 34K^2$ parameters, while the improved version has $4NK + 34K + 2K^2$ parameters–$32K^2 - 34K$ fewer than the original.

2.4 Improved Multiscale Neural Network

Referencing the VGG-16 architecture, this paper designs a multiscale convolutional neural network based on the DS-Inception block, termed DSR-Inception (see Figure 1 [Figure 1: see original paper]). The network comprises pooling modules and convolution modules containing DS-Inception blocks, ReLU activation functions, and batch normalization. Replacing VGG-16' s five traditional convolutional modules with the proposed modules substantially reduces model parameters. ReLU activation is assigned to all convolutional and fully connected layers, while the Sigmoid function is applied to the final output layer for classification. Each convolutional layer is followed by a 2×2 max-pooling layer with stride 2 to aggregate transmitted information. Filter counts increase with network depth, enabling extraction of finer details from input images.

The network model' s processing flow is illustrated in Figure 2 [Figure 2: see original paper]. A $224 \times 224 \times 3$ face image enters through two branches. One branch passes through the DS-Inception block, where each branch uses 16 convolution kernels, outputting a $224 \times 224 \times 64$ feature map. The other branch adjusts channel count via a residual block. Feature maps from both branches are summed element-wise before passing through ReLU activation, batch normalization, and pooling layers. After five such operations, face features are fully extracted, yielding a $7 \times 7 \times 1024$ feature map. This map then undergoes global average pooling and dropout before final classification by a sigmoid classifier.

The designed multiscale network incorporates convolution kernels of various sizes to extract multi-dimensional features, solving the low accuracy problem associated with single-scale networks. Additionally, the residual structure accelerates training convergence and prevents overfitting.

2.5 Joint Supervised Loss Function

Classification optimization essentially minimizes an objective function. Softmax loss is commonly used for multi-class image classification, defined by Equation (19):

$$L_S = -\frac{1}{m} \sum_{i=1}^{m} \log \frac{e^{{W_{y_i}}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} \tag{19}$$

While softmax loss offers good differentiability, it lacks discriminative power. Face recognition tasks exhibit high inter-class feature redundancy, potentially increasing variation between faces of the same person.

Center loss function [31] is a clustering algorithm that drives each class toward a center point, effectively imposing strong constraints per class. It is defined by Equation (20):

$$L_C = \frac{1}{2} \sum_{i=1}^{m} \|x_i - c_{y_i}\|_2^2 \tag{20}$$

The joint supervised loss function combines softmax and center loss, simultaneously optimizing classification accuracy and feature center clustering for face recognition. The schematic is shown in Appendix 3.

The softmax loss function maximizes face classification accuracy by minimizing the difference between the model' s output probability distribution and true labels. Conversely, the center loss function clusters feature vectors of the same category near a center point while separating different categories to enhance identity information identification. The center loss incorporates a weight parameter λ that regulates center point influence and update speed. The joint supervised loss function is shown in Equation (21):

$$L = L_S + \lambda L_C \tag{21}$$

where m and n represent the number of samples and categories, x_i is the image feature, y_i is the category label, W_j is the fully connected layer weight, b is the bias, c_{y_i} is the class center, λ is the balance coefficient with $\lambda \in (0, 0.1)$, and $\lambda = 0.01$ in this paper.

3.1 Experimental Platform

Model training is conducted in a GPU environment with configuration details shown in Table 1 .

3.2 Dataset

The pre-training dataset is the CelebA face dataset, which provides face images with attribute classifications and facial landmark annotations. Sample face images from this dataset are shown in Appendix 4.

The homemade miner face dataset is designated as the MF dataset, containing 40 individuals with 21 images each—7 images without coal ash occlusion, 7 with light occlusion, and 7 with heavy occlusion—totaling 840 images. Table 2 shows sample face data for a single person.

3.3 Contrast Experiment

This study employs stochastic gradient descent (SGD) as the training optimizer, with hyperparameters set as follows: batch size 64, initial learning rate $\alpha_0 = 0.01$, decay index $\beta = 0.05$, and epoch count 150. Inception-v1, VGG-16, and ResNet-18 serve as comparison models, trained on the CelebA dataset. Training accuracy and loss curves are shown in Appendix 5, demonstrating that the proposed DSR-Inception network exhibits superior convergence speed and accuracy compared to other models.

Different transfer learning scheme designs are presented in Table 3 . Experimental results in Appendix 6 indicate optimal transfer learning performance when Conv1-Conv3 module weights are frozen and remaining layers are retrained.

The CelebA dataset is used for pre-training (14,000 training images, 4,000 test images). Comparative results appear in Table 4 , showing that the proposed model achieves higher accuracy and recall than other networks. The proposed model size is only 46.56 MB, with average testing time of 258 ms per face image.

The MF dataset is used for fine-tuning, with comparative experiments testing the improved loss function. Results in Appendix 7 show that across 20 experimental runs, the joint supervised loss function improves F1-score by approximately 0.5–1.5%, confirming its effectiveness.

Table 5 compares the proposed model with recent mainstream methods on the CelebA dataset. The proposed method outperforms most existing approaches, demonstrating that fusing multi-scale facial features effectively improves occluded face detection accuracy.

4. Application of Proposed Approach in Underground Coal Mine

To further validate feasibility, the face recognition system is deployed in a tunneling tunnel. The KBA18W mining monitoring camera captures images transmitted via underground wireless routers, with LED light sources as auxiliary lighting. Field equipment setup is shown in Figure 3(a) [Figure 3: see original paper].

Before testing, six staff members' facial data are registered in the system, with the display interface shown in Figure 3(b) [Figure 3: see original paper].

A 60-hour working period validates system feasibility, with recognition results summarized in Table 6 . The system achieves 91.34% average recognition accuracy in coal mine excavation tunnels with a miss detection rate below 5%, meeting design requirements.

5. Conclusions and Future Work

This paper proposes an improved face recognition method for underground coal mines based on transfer learning to address random coal dust occlusion. A novel DS-Inception block reduces model parameters, and a joint supervised loss function based on center and softmax loss adapts to face recognition classification tasks. Compared with classical models (Inception-v1, VGG-16, ResNet-18), the proposed multiscale neural network achieves accuracy, recall, and F1-score of 97.26%, 94.17%, and 95.42%, respectively. To better adapt to coal mine environments, a miner face dataset is created for transfer model fine-tuning, with the joint supervised loss improving accuracy by approximately 0.5–1.5%. Additionally, the face recognition system achieves 91.34% average accuracy in excavation tunnels with under 5% miss detection rate.

While this work validates the proposed method' s effectiveness in underground coal mines, it has not yet addressed face recognition under large-angle posture variations. Future research will investigate face recognition under diverse postures.

Acknowledgements

The research leading to these results received funding from the Norwegian Financial Mechanism 2014-2021 under Project Contract No. 2020/37/K/ST8/02748.

Conflict of Interests

The authors declare no conflict of interests regarding the publication of this article.

Data Availability

All data producing the results in this work can be requested from the corresponding author.

References

[References remain unchanged from original]

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