# Postprint of Image Distortion Type Determination Algorithm Based on Gabor Wavelet and CNN

**Authors:** Pengcheng Li, Wu Tao, ZHANG Shanqing

**Date:** 2018-06-19T00:00:00+00:00

## Abstract

A novel distortion type determination algorithm based on Gabor wavelets and CNN (convolutional neural network) is proposed for image distortion classification. The algorithm first leverages the favorable characteristics of Gabor wavelets for coarse feature extraction from images, and subsequently employs an improved CNN to further extract key features. The algorithmic procedure comprises: first, preprocessing the images (including label assignment, sample balancing, and sample augmentation); second, performing an 8-directional Gabor wavelet transform on the preprocessed images and superimposing subbands from different directions to constitute input samples; and finally, training the samples using a self-designed CNN and Softmax classifier, wherein convolution kernel parameters are optimized via stochastic gradient descent and backpropagation to obtain the final model. Experiments for distortion type determination conducted with the trained model achieve a classification accuracy of 95.62% on the LIVE standard image database, demonstrating that the proposed algorithm exhibits high accuracy and robustness.

## Full Text

## Preamble

### Image Distortion Type Classification Algorithm Based on Gabor Wavelet and CNN

*Li Pengcheng, Wu Tao, Zhang Shanqing†*

(School of Computer Science & Technology, Hangzhou Dianzi University, Hangzhou 310018, China)

**Abstract:** For image distortion classification, this paper proposes a novel algorithm based on Gabor wavelet and CNN (Convolutional Neural Network). The algorithm first leverages the desirable properties of Gabor wavelet for coarse feature extraction from images, then employs an improved CNN to further extract key features. The main steps include: (1) Image preprocessing (including label assignment, sample balancing, and sample augmentation); (2) Performing 8-directional Gabor wavelet transform on preprocessed images and superimposing sub-bands from different directions to construct input samples; (3) Training the samples using a self-designed CNN and Softmax classifier, with convolution kernel parameters optimized during training via stochastic gradient descent and backpropagation to obtain the final model. Experiments on the LIVE standard image database demonstrate a classification accuracy of 95.62%, indicating that the proposed algorithm achieves high accuracy and robustness.

**Keywords:** convolutional neural network; Gabor wavelet; image distortion type; feature learning

## 0 Introduction

Images inevitably suffer from various types and degrees of distortion during acquisition, compression, processing, and transmission. These distortions not only degrade visual experience but can also cause significant losses. Consequently, research on image quality assessment has attracted increasing attention, particularly regarding the classification of different distortion types and the development of evaluation algorithms tailored to specific distortions.

Currently, several algorithms target specific distortion types, such as those for blurred images and noisy images. Before applying these algorithms, distortion type classification must be performed first. Since different distortion types exhibit distinct features—some easily distinguishable while others readily confused—distortion type classification remains a challenging problem. Existing solutions can be broadly categorized into two classes: The first extracts spatial or frequency domain features and applies traditional machine learning methods for classification. The second adopts deep learning approaches.

The most common methods in the first category utilize NSS (Natural Scene Statistics) features. NSS characteristics of natural images can be extracted in the DCT or wavelet domain, described by appropriate mathematical statistical models whose parameters vary with different distortion types and levels. Representative NSS-based methods include BIQI, DIIVINE, and BRISQUE. BIQI extracts statistical features from wavelet coefficients of distorted images and employs traditional machine learning for distortion classification. DIIVINE performs multi-scale wavelet decomposition on distorted images, fits mixture Gaussian model parameters using NSS properties as image features, and uses these features for distortion type identification. BRISQUE extracts spatial NSS features and utilizes an SVM classifier for distortion type classification.

The second category comprises deep learning-based methods. Compared with traditional machine learning, deep learning can achieve higher evaluation accuracy. For instance, Hou et al. applied deep belief networks with multiple hidden layers to image quality assessment. Rather than using regression to predict image quality, they adopted a deep learning classification approach by first categorizing image quality into classes such as "excellent," "good," "fair," "poor," and "bad," then computing posterior probabilities for each class and combining them as the final quality score. This work demonstrated the feasibility of training deep networks via distortion degree classification. In another study, Wu et al. partitioned images of different distortion types into patches, applied local contrast normalization, and used a trained CNN for distortion type classification. Experiments on the LIVE image database showed superior performance compared to BIQI, DIIVINE, and BRISQUE, though accuracy still had considerable room for improvement. Generally, deep learning methods achieve higher accuracy, but existing algorithms suffer from large intra-class variation exceeding inter-class variation due to wide distortion degree ranges within each type, leading to significant errors.

To address these issues, this paper proposes an image distortion type classification algorithm based on Gabor wavelet and CNN. Unlike previous work, our algorithm not only categorizes distortion types but also sub-classifies each type by distortion degree. To better perform coarse feature extraction, Gabor wavelet transform is applied before training. Additionally, different CNN structures are employed for training. Experiments demonstrate that the improved network converges faster and yields more accurate classification results.

## 1 Classification Algorithm

The distortion type classification algorithm consists of four main steps: (a) Preprocessing, including sample augmentation, sample balancing, and label assignment; (b) 8-directional Gabor wavelet transform for signal decomposition, with sub-bands superimposed as input samples; (c) Training the superimposed samples using a CNN neural network and Softmax classifier; (d) Using the trained model to determine image distortion types. The algorithm flow is shown in [Figure 1: see original paper].

### 1.1 Sample Preprocessing

**1.1.1 Sample Augmentation** Before training, to maximize sample quantity, all samples undergo patch partitioning. Specifically, each image is divided into overlapping N×N patches, with each patch serving as a sample. The values of N and the overlapping stride are determined experimentally. To ensure validity, images with background blur also undergo cropping processing, as shown in [Figure 2: see original paper].

**1.1.2 Sample Balancing** Since different classes may have uneven sample distributions, leading to inaccurate classification results, affine transformations are applied to some image patches to balance the dataset. The balancing process proceeds as follows:

a) Calculate the ratio between sample quantities of different classes: $\rho = \frac{N_a}{N_b}$, where $N_a$ and $N_b$ are the sample counts for class A and class B, respectively. When $\rho$ falls within a reasonable range, samples are considered balanced. The reasonable interval is $\rho \in [\sqrt{2}/2, \sqrt{2}]$.

b) If $\rho$ is outside the reasonable interval, apply affine transformations (rotation or scaling) to samples until $\rho$ falls within the reasonable range.

**1.1.3 Sample Label Assignment** During label assignment, besides categorizing by distortion type, each distortion type is further sub-classified by distortion degree. For example, for images in the LIVE database with Gaussian blur (Gblur) and white noise (WN) distortions, the labeling process is: First, classify by distortion type into two categories; then normalize the DMOS (Differential Mean Opinion Score) values to [0,1] and equally divide them into three intervals; finally, sub-classify each image according to its DMOS interval, as detailed in . This sub-classification reduces intra-class differences, preventing network confusion during training and improving accuracy.

### 1.2 Gabor Wavelet Transform

Before formal training, Gabor wavelet transform is applied to samples. Unlike other wavelets, Gabor wavelet captures both frequency and spatial domain information. The selection rationale includes: (a) Gabor wavelet offers multi-scale, multi-directional selectivity, is sensitive to edges but insensitive to illumination, thus providing good illumination invariance; (b) The 2D Gabor wavelet kernel exhibits excellent spatial locality, capable of extracting multi-scale, multi-directional local structural features similar to human visual cells; (c) Gabor wavelet transform for feature extraction requires less data, offers faster processing, and demonstrates good robustness. Additionally, NSS features in the wavelet domain are sensitive to image distortion types.

[Figure 3: see original paper] shows histogram representations of different distortion types after Gabor wavelet sub-band superposition, with normalized data. The histogram shapes vary significantly across distortion types, indicating that distortion types can be distinguished using Gabor wavelet transform.

The Gabor wavelet transform is represented by Equation (2):

$$\phi'_{p,q}(x,y) = \sum_{s}\sum_{t} I(x-s, y-t)\phi_{p,q}(s,t)$$

where $I(x,y)$ is the original image, $x$ and $y$ are pixel coordinates, $s$ and $t$ are filter element positions, $p \in \{0,1,\dots,P-1\}$ and $q \in \{0,1,\dots,Q-1\}$ represent scale and orientation parameters. $\phi_{p,q}(x,y)$ is the complex conjugate of the

Gabor wavelet transform function, whose definition and solution can be found in references. After convolution with 8 directional kernels, 8 sub-band images are produced, then superimposed via Equation (3) to obtain training samples:

$$S = \frac{1}{Q}\sum_{i=0}^{Q-1} I_i$$

where $I_i$ is the $i$-th directional sub-band image.

### 1.3 CNN Architecture Details

**1.3.1 Activation Function Selection** Since Gabor wavelet processing generates many sparse features, the LReLU (leaky rectified linear) activation function is employed to facilitate data representation and accelerate convergence, as shown in Equation (4). This is an improved version of the ReLU function (when coefficient $a = 0$, it becomes ReLU). The modification not only benefits sparse data processing but also addresses the potential issue of some neurons remaining permanently inactive in ReLU.

$$h(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{if } x \leq 0 \end{cases}$$

where $W$ represents weights, $x$ represents features, and $a$ is a small constant near 0, determined experimentally.

**1.3.2 Convolution Kernel Determination** Convolution kernel weights are determined by minimizing the network loss function. The loss function $J(W, b)$ uses squared loss with regularization to prevent overfitting:

$$J(W, b) = \frac{1}{m}\sum_{i=1}^{m}\left[\frac{1}{2}(L(W, b; x^{(i)}, y^{(i)}))^2\right] + R(h)$$

$$R(h) = \frac{\lambda}{2}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^{(l)})^2$$

where $m$ is the sample size, $y^{(i)}$ is the predicted label, $\lambda$ is the weight decay parameter, $R(h)$ is the regularization term, $l$ is the hidden layer index, $j$ is the neuron index in each hidden layer, and $i$ indicates the weight index connected to each neuron.

To minimize the cost function, stochastic gradient descent and backpropagation are employed for parameter updates. Parameters $W$ and $b$ are updated according to Equation (7):

$$\begin{cases} W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W,b) \\ b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W,b) \end{cases}$$

where $\alpha$ is the learning rate controlling gradient descent step size. To improve convergence, fixed step size is replaced with dynamic step size as follows:

$$\alpha_t = \begin{cases} \alpha_{t-1}^{*} & \text{if } J_t \geq J_{t-1} \\ \alpha_{t-1} & \text{if } J_t < J_{t-1} \end{cases}$$

where superscript $t$ denotes iteration number and $J_t$ is the cost function value. Additionally, to prevent overfitting, the Dropout method is applied in fully connected layers. This method, proposed by Hinton et al., randomly ignores the influence of some hidden neurons during training with a certain probability.

**1.3.3 Network Structure** The network structure is shown in [Figure 4: see original paper]. The convolutional network contains 11 layers: input layer, 4 convolutional layers, 4 pooling layers, fully connected layer, and output layer. The first convolutional layer uses 32 single-channel 5×5 filters, followed by Max-Pooling. The second convolutional layer uses 64 3×3 32-channel filters, with identical pooling. The third convolutional layer uses 32 3×3 64-channel filters, with the same pooling method. The fourth layer applies 128 5×5 32-dimensional filters, again followed by pooling. The fully connected layer reshapes neurons from the previous pooling layer into a 1D vector, computes with 2048 fully connected neurons to obtain a 2048-dimensional feature vector, includes a Dropout layer that filters the 2048-dimensional features to produce the final feature vector. The output layer classifies using a Softmax classifier.

## 2 Experiments

The experimental environment consists of an Intel® Core™ i5 CPU at 2.80GHz, 8GB RAM, 64-bit Windows 7 Ultimate OS, using Python 3.6.1 with TensorFlow 1.2.1 framework. The LIVE database is used with distortion types: Gaussian blur (Gblur), JPEG compression (JPEG), white noise (WN), JPEG2000 compression (JP2K), and fast fading Rayleigh channel distortion (FF), with sample counts of 145, 204, 145, 198, and 145 images respectively. Before training, all samples are randomly divided into training (60%), validation (20%), and test (20%) sets.

### 2.2.1 Extreme Sample Classification Analysis

To verify robustness, [Figure 5: see original paper] shows four images of different distortion types with normalized DMOS values. lists classification results from different algorithms. The DMOS values indicate these images are near distortion-free, making visual distinction difficult. However, the proposed algorithm accurately classifies all distortion degrees and provides approximate distortion levels. In contrast, other algorithms exhibit some misclassifications on these challenging samples.

#### 2.2.2 Overall Database Comparison

To demonstrate accuracy and robustness, [Figure 6: see original paper] presents the confusion matrix for all images in the LIVE database. Matrix values indicate confusion degrees between distortion types, where the value at row $i$, column $j$ represents the probability of classifying type $i$ distortion as type $j$. The matrix shows JP2K and FF distortions are somewhat confused, JPEG and JP2K have slight confusion, while WN and Gblur show almost no confusion with other types. Overall, confusion probabilities are low, indicating the algorithm is particularly suitable for JPEG, WN, and Gblur distortion classification.

compares the proposed algorithm' s accuracy with other methods on the LIVE database, including per-distortion and overall accuracy. Except for WN distortion (where accuracy is slightly lower than some methods), the proposed algorithm outperforms others for all distortion types. The overall accuracy surpasses existing methods.

### 2.3 Parameter Comparison Experiments

Numerous parameters exist in the algorithm. Cross-validation experiments on the LIVE database were conducted to determine optimal values.

**2.3.1 Patch Size Selection** Using the control variable method, experiments were performed with different patch sizes. To avoid sample imbalance, overlapping patch sampling was adopted. As shown in , 64×64 patches yield the highest overall accuracy.

**2.3.2 Sampling Stride** With patch size fixed at 64×64, experiments were conducted to determine the optimal sampling stride. [Figure 7: see original paper] shows stride versus overall accuracy, demonstrating that a stride of 32 achieves the highest accuracy.

**2.3.3 Convolution Kernel Size** To determine kernel sizes, the control variable method was applied again. With all other conditions fixed, all kernels were initially set to 3×3, then each layer' s kernel size was varied sequentially to find optimal dimensions. Results in show highest accuracy is achieved with: first kernel 5×5, second and third kernels 3×3, and fourth kernel 5×5.

**2.3.4 LReLU Activation Coefficient** To determine the constant $a$ in the activation function, experiments were conducted while keeping other parameters constant. [Figure 8: see original paper] shows the relationship between $a$ values and overall accuracy. The highest overall accuracy occurs at $a = 0.004$.

Runtime comparison in shows the proposed algorithm, while not as fast as BIQI or BRISQUE, achieves near real-time performance.

## 3 Conclusion

To address image distortion type classification using deep learning and solve the problem of intra-class variation exceeding inter-class variation in previous methods, this paper proposes a Gabor wavelet and CNN-based algorithm. The method first applies multi-directional Gabor wavelet transform, superimposes sub-bands as samples, trains them using a self-designed CNN to obtain the final model, and then performs distortion type classification. Experiments on the LIVE database demonstrate high recognition rates for the proposed approach.

## References

[1] Wang Zhengzi, Xie Zhihua, He Cuiqun. A fast quality assessment of image blur based on sharpness [C]// Proc of International Congress on Image and Signal Processing. 2010: 2302-2306.

[2] Li Leida, Lin Weisi, Wang Xuesong, et al. No-reference image blur assessment based on discrete orthogonal moment [J]. IEEE Trans on Cybernetics, 2016, 46 (1): 39-50.

[3] Cui Guangmang, Feng Huajun, Xu Zhihai, et al. No-reference image quality assessment based on CSF and affine reconstruction model [J]. Journal of Zhejiang University: Engineering Science, 2016, 50 (1): 144-150.

[4] Li Congli, Yang Xiushun, Chen Wenbing, et al. Study on the IQA method for polarization image based on degree of noise pollution [C]// Proc of International Conference on Information and Automation. 2009.

[5] Moorthy A K, Bovik A C. A two-Step framework for constructing blind image quality indices [J]. IEEE Signal Processing Letters, 2010, 17 (5).

[6] Moorthy A K, Bovik A C. Blind image quality assessment: from natural scene statistics to perceptual quality [J]. IEEE Trans on Image Processing, 2011, 20 (12): 3350-3364.

[7] Mittal A, Moorthy A K, Bovik A C. No-reference image quality assessment in the spatial domain [J]. IEEE Trans on Image Processing, 2012, 21 (12).

[8] Wu Meiyin, Chen Lin, Tian Jing. Video image distortion detection and classification based on convolution neural network [J]. Application Research of Computers, 2016, 33 (9): 2827-2830.

[9] Ruderman D L. The statistics of natural images [J]. Network Computation in Neural Systems, 1994, 5 (4): 517-548.

[10] Hou Weilong, Gao Xinbo, Tao Dacheng, et al. Blind image quality assessment via deep learning [J]. IEEE Trans Neural Networks & Learning Systems, 2017, 26 (6): 1275-1286.

[11] Zhang Gang, Ma Zongmin. A texture feature extraction method using Gabor wavelet [J]. Journal of Image and Graphics, 2010, 15 (2): 247-254.

[12] Mass A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models [C]// Proc of International Conference on Machine Learning. 2013.

[13] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines [C]// Proc of International Conference on International Conference on Machine Learning. Omnipress. 2010: 807-814.

[14] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. Computer Science, 2012, 3 (4): 212-223.

[15] Sheikh H R, Sabir M F, Bovik A C. A statistical evaluation of recent full reference image quality assessment algorithms [J]. IEEE Trans on Image Processing, 2006, 15 (11): 3440-3451.

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