Hindawi
Security and Communication Networks
Volume 2021, Article ID 8297244, 11 pages
https://doi.org/10.1155/2021/8297244

Research Article

Perceptual Image Hashing Based on Multitask Neural Network

Cheng Xiong, Enli Liu, Xinran Li, Heng Yao, Lei Zhang, and Chuan Qin

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Correspondence should be addressed to Chuan Qin; qin@usst.edu.cn

Received 2 November 2021; Accepted 2 December 2021; Published 18 December 2021

Academic Editor: Beijing Chen

With the advent of the era of multimedia and in-depth development, the whole human society has been produced and spread a huge amount of image data, but at the same time, in view of the digital image and tamper with the attack of piracy phenomenon also more and more serious, malicious attacks will produce serious social, military, and political influence, therefore, to protect the authenticity of the original image content, which is also more and more important. In order to further improve the performance of image hashing and enhance the protection of image data, we proposed an end-to-end dual-branch multitask neural network based on VGG-19 to produce a perceptual hash sequence and used prepart of network of pretrained VGG-19 model to extract image features, and then, the image features are transformed into a hash sequence through a convolutional and fully connected network. At the same time, in order to enhance the function of the network and improve the adaptability of the proposed network to using scenarios, the rest part of the network layer of the VGG-19 model was used as another branch for image classification, so as to realize the multitask characteristics of the network. Through the experiment of the testing set, the network can not only resist many kinds of attack operations (content retention operations), but also realize accurate classification about the image, and has a satisfactory tampering detection ability.

1. Introduction

Since the beginning of this century, the technologies of Internet and multimedia develop rapidly, and information interaction mode of people has been transformed from text message to multidirectional fusion presentation of text message, image, and video information. With the wide application of powerful image editing tools, massive digital images are easily tampered; thus, the protection of the real content of the image is increasingly important. Due to the influence of technology, equipment, time, and other factors, an image is often distributed without any protection after it is produced, which makes the image more vulnerable to piracy, tamper, and other attack operations. In order to deal with a variety of malicious attacks, image hashing technology can be used to generate a unique and unidirectional perceptual image hash sequence for the original image. Image hashing, also known as "image fingerprint," verifies image piracy or tamper by comparing the similarity of hash sequences. Imagine a scenario that there is an image that is so important for the owner. But he does not know if his image was tampered or pirated, and it is so difficult for him to check in the image dataset, which has so many images. So, the image hashing can help him to find the similar images quickly by comparing the similarity of hash sequences effectively and judge whether they are tampered or piracy images to protect the copyright.

With the in-depth research and development of image hashing schemes, a variety of schemes have been proposed by researchers according to various requirements of multimedia security. Generally speaking, a typical perceptual image hashing scheme has the following three properties: (1) *perceptual robustness*—the generated hash sequence of the image without changing the visual content after the content retention operations should be similar or the same as the hash sequence of the original image; (2) *discriminative/anticollision*—two hash sequences correspond to two completely different images should be completely different; and (3) *security*—the hash sequence needs to have key dependence to ensure the security of scheme; that is, the hash

sequence generated with the wrong key is completely different from the sequence generated with the right key. The perceptual image hashing scheme includes three main stages: preprocessing, feature extraction, and hash generation. According to the methods of feature extraction, perceptual image hashing schemes can be divided into four main types: methods based on spatial domain, transform domain, dimensionality reduction, and deep learning.

In the methods based on the spatial domain, Schneider and Chang [1] used the histogram features of the image to generate hash sequence, which opened up the research in the field of image hashing. Yan et al. [2] proposed adaptive local feature extraction technology to obtain the location information of features and achieve image tampering positioning. In order to resist the attack of image rotation from any Angle, some scholars [3] used MDS (Multidimensional Scaling Technology) [4] on the basis of ring-based coding scheme, ring division, and invariant vector distance [5], and experimental results show that the hash sequences generated by this scheme are robust and unique for common image content retention operations. The scheme in [6] used the relationship between local feature points to overcome the problem, which feature point distribution is ignored. Qin et al. [7] proposed a robust image hashing scheme based on perceived texture and structural features; furthermore, local texture features and color vector angle features were considered simultaneously in [8]. Shen et al. [9] extracted the color opposition component from the secondary image and applied quadtree decomposition to connect the generated color feature vector with the structural feature vector and then combined with pseudo-random key scrambling to generate the final hash sequence. For the content of image in screen, the scheme in [10] extracted the maximum gradient and the corresponding direction information from R, G, and B color components and counted the relevant data to construct the image hash sequence. There are also some research studies about retrieval, and the scheme in [11] proposed a content-based image retrieval scheme in multiuser scenarios, which used Euclidean distance comparison technology to sort the similarity of image feature vectors and return top-K results to achieve the retrieval purpose. Li et al. [12] proposed an encrypted image retrieval system supporting multiple keys with edge computing based on local sensitive hashing, secure neighbor, and proxy re-encryption technologies, which improved the efficiency and accuracy of image retrieval.

In the methods based on the transform domain, some early schemes used DCT [13] (discrete cosine transform) and DFT (discrete Fourier transform) to design schemes. From the perspective of human visual characteristics, Watson was used to adjust the corresponding frequency domain coefficients to improve the robustness and discrimination of the scheme in [14]. Some scholars also used the DFT [15] to extract robust frequency features in the secondary image, and nonuniform sampling was used to combine with low and intermediate frequency components to obtain a secure hash sequence. In the work of [16], the CSLBP (center symmetric local binary pattern) was applied to DWT (discrete wavelet transform) to generate compact image hashing. A geometric invariant vector distance method based on both the spatial domain and the frequency domain was proposed in [17]. In the dimensional-reduction method, NMF (non-negative matrix factorization), PCA (principal component analysis), and SVD (singular value decomposition) often occur as important steps. In [18], hash sequences were generated by combining BTC (block truncation coding) with PCA. A scheme based on low-rank sparse decomposition was proposed in [19].

In recent years, with the continuous improvement of GPU performance, and continuous development of deep learning, so many researchers have used deep neural networks to achieve image retrieval and the performance is much better than the traditional image retrieval schemes [20–23]. But, generally speaking, there are relatively few research studies on perceptual robust hashing used deep neural networks. In the methods based on deep neural networks, the scheme in [24] proposed a robust image hashing scheme, which is based on deep learning. And this work was an early scheme to perceptual image hashing with deep neural networks, and it has better performance than traditional schemes. In the work of [24], researchers have used pretrained DAE (auto encoder denoising) to enhance robustness and used fine-tuning to improve the accuracy of image detection. In [25], an image hashing scheme based on CNN (convolutional neural network) with multiple constraints was proposed, and the experimental results showed that it can obtain a good balance between robustness and discrimination. In order to strengthen the functional properties of the neural network and make more efficient and reasonable use of existing computing resources, there are also some other schemes about deep learning in the field of image authentication. The scheme in [26] introduced two subnetworks to improve the BusterNet for image copy-move forgery localization with source/target region distinguishment, and these subnetworks were the copy-move similarity detection network (CMSDNet) and the source/target region distinguishment network (STRDNet). The face detection is also important. Reference [27] used RGB and YCbCr color spaces, and introduced the convolutional block attention module and multilayer feature aggregation module into the Xception model to achieve better performance for detecting postprocessed face images.

We proposed a perceptual image hashing scheme based on the multitask neural network. The mainly innovation and contribution are as follows:

(1) Efficient end-to-end framework. Instead of the traditional strong explanatory method with low efficiency and weak generalization ability, an advanced and efficient deep learning method is adopted to collect image features and generate hash sequence. Based on the excellent performance of the convolutional network and fully connected network, the end-to-end hash sequence generated framework is realized by integrating feature extractor and hash generator.

(2) Multitask intensifies the applicability of network model to multiple scenarios. In order to improve the

applicability of the neural network model to multiple scenarios and use an excellent pretrained model, based on the pretrained model of VGG-19 neural network, we added a dual-branch layer after the feature extractor, so as to achieve the purpose of multitask.

In Section 2, the structure of the proposed network, loss function, and other related contents will be introduced. In Section 3, the experimental results of the proposed neural network based on a specific training strategy will be introduced. The advantages of our scheme will be summarized in Section 4.

2. Proposed Scheme

At present, the field of deep learning is developing rapidly, a large number of network structures have emerged, and the use of pretrained models is becoming more and more systematic. Many network structures are used for different tasks. On the basis of realizing the image hash process, the proposed scheme has added the function of image classification, so that the final neural network not only can generate the hash sequence of the image but also has function in the image classification, which is adapted to more application scenarios.

2.1. Network Structure. In the structure of the proposed neural network, we comprehensively considered the task requirements on the number of network layers and network structure, and selected VGG-19 as the basic structure to use. We first use part of VGG-19 network layer as the image feature extractor. Note that, because the function of image classification should be considered, this part of VGG-19 does not participate in the parameter updating of the training in proposed neural network, but uses the fixed parameters. The main structure of the proposed neural network is described in Figure 1. The feature extractor is connected with a small convolutional network of six layers and constructed hash generation network, which is named Branch I, and the connected point of this generated network is named branch point. The proposed network has selected the output in the second pooling of the VGG-19 network as the branch point through testing in the network training process. Specific test and evaluation will be discussed in Subsection 2.2. At the same time, the rest part (Branch II) of the network layer in VGG-19 also is remained and connected with the feature extractor through branch point to realize the function of image classification. Branch II was composed of twelve convolutional layers and three FC (Fully connected) layers, and the activation function was ReLU. So that, due to branch point, multitask is achieved and constructed the dual-branch neural network.

During the processing stage of the image hashing neural network, the feature extractor is used to collect features of the image. Then, the features are input into the small convolutional network to generate the hash sequence, and the small convolutional network is mainly composed of four blocks (convolutional layer + BN + ReLU) and two FC layers. Image features and intermediate features are so important, and the convolutional layer is a useful filter to extract them. Thus, we used four convolutional blocks in the hashing neural network to further extract useful features, and two FC layers were used to compress and transform the features into a hash sequence. As for the other branch, it is used to process features, which can classify images and realize the multitask capability. Due to the use of FC and average pool layers, the input size of image was limited, and the size was 128×128.

2.2. Loss Function. In order to measure the similarity of generated hash sequences, we use MSE (mean square error) as the measurement tool to calculate the hash distance among original image, similar images, and different images. However, due to the large numerical span of the generated hash sequence, direct use of MSE will make the network convergence unstable, so the activation function named Sigmoid is used to normalize the values obtained by MSE.

We define that there are n similar images and n different images (the specific composition of the dataset will be introduced in Subsection 2.3, Subsection 3.1, and Subsection 3.2), so that, in the situation of including the original image, $2n + 1$ hash sequences are generated, which is given in the following equation:

$$H(\chi), H\left(\widehat{\chi}_{[1]}\right), \ldots, H\left(\widehat{\chi}_{[n]}\right), H\left(\kappa_{[1]}\right), \ldots, H\left(\kappa_{[n]}\right), \quad (1)$$

where $H(\cdot)$ represents the whole end-to-end hashing neural network, χ is the original image, $\widehat{\chi}_{[i]}$ is similar image ($i = 1, 2, \ldots, n$), and $\kappa_{[i]}$ represents the different image ($i = 1, 2, \ldots, n$). The hash sequences of images are generated, and the MSE is used to measure the distance of sequences, which are given as

$$\begin{aligned} \varphi_{sh}^{(i)} &= \frac{1}{l}\sum_{k=1}^{l}\left|H(\chi)_k - H\left(\widehat{\chi}_{[i]}\right)_k\right|^2, \\ \varphi_{dh}^{(i)} &= \frac{1}{l}\sum_{k=1}^{l}\left|H(\chi)_k - H\left(\kappa_{[i]}\right)_k\right|^2, \end{aligned} \quad (2)$$

where $|\cdot|^2$ represents the Euclidean norm, l is the length of hash sequence, and $\varphi_{sh}^{(i)}$ and $\varphi_{dh}^{(i)}$ represent the hash distance of similar image pair and different image pair, respectively. The hash distance between original and similar image, and between similar and similar images should be small; on the other hand, the hash distance between original and different images should be large. So, as for the proposed scheme, $\varphi_{sh}^{(i)}$ should be as small as possible, and $\varphi_{dh}^{(i)}$ should be as larger as possible. In order to obtain useful loss value, the Sigmoid is used to normalize the value of MSE to [0.5, 1], as shown in the following equation:

$$S(x) = \frac{1}{1 + e^{-x}}, \quad (3)$$

where $S(\cdot)$ is the function of Sigmoid, and the whole loss function is

$$\begin{aligned} &\min \Gamma = \alpha_1 \frac{1}{n}\sum_{i=1}^{n} S\left(\varphi_{sh}^{(i)}\right) - \alpha_2 \frac{1}{n}\sum_{i=1}^{n} S\left(\varphi_{dh}^{(i)}\right), \\ &\text{s.t. } \alpha_1, \alpha_2 \in [0, 1], \end{aligned} \quad (4)$$

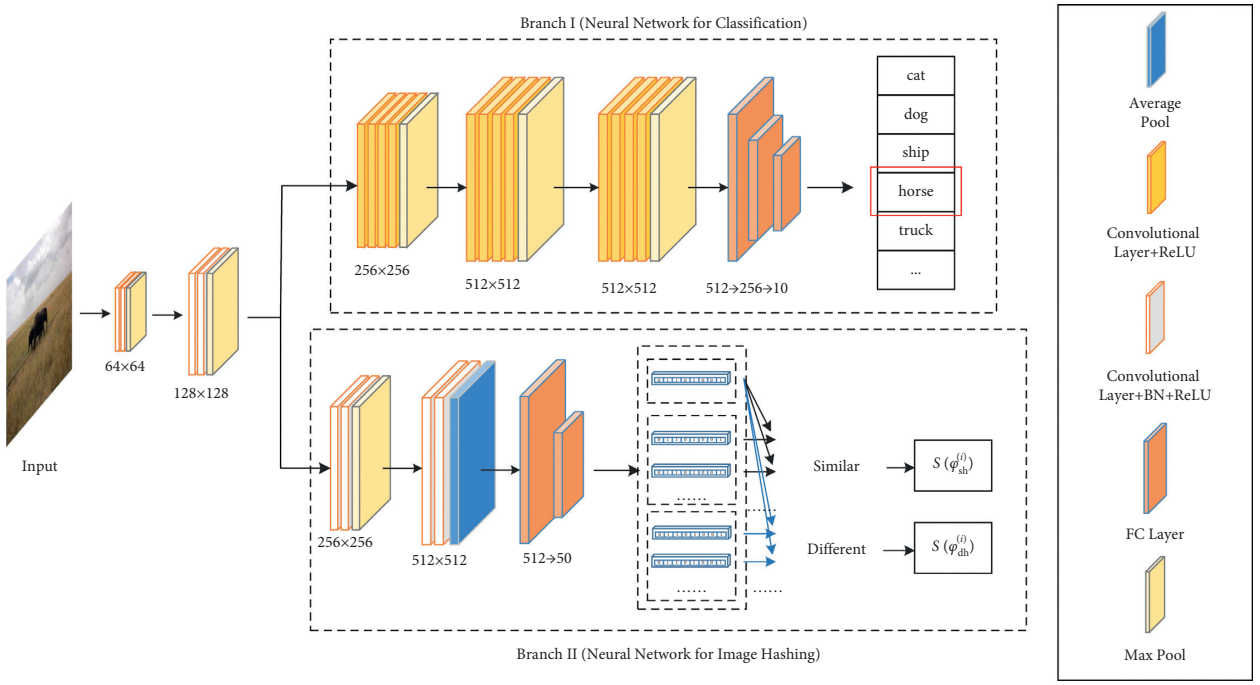


Figure 1: Structure of the dual-branch multitask neural network.

where Γ represents the whole loss function, and α_1 and α_2 both represent the super parameters, which are used to adjust the loss function. As for the pretrained model of VGG-19, the CEL (cross-entropy loss) function is used during the training. Denote a sample's tag *target* as $\{t_1, t_2, ..., t_{10}\}$, the predicted output of the model is $\{o_1, o_2... o_{10}]$, and then the value of CEL of the network is

$$L = -\sum_{i=1}^{10} t_i \log(o_i). \tag{5}$$

In particular, the following is used to measure and decide the best branch point among these max pooling layers:

$$\varsigma = \max_{j=1,2,\ldots,m_v} \left\{ \frac{1}{n_v} \sum_{i=1}^{n_v} S\left(\varphi_{sh}^{(i)}\right) - \frac{1}{n_v} \sum_{i=1}^{n_v} S\left(\varphi_{dh}^{(i)}\right) \right\}_j, \tag{6}$$

where $\varphi_{sh}^{(i)}$ and $\varphi_{dh}^{(i)}$ represent the hash distance of perceptual similar pair and different image pair, respectively. The m_v is the quantity of batches in the test image set, and n_v is the quantity of similar images or different images in one batch. The smaller the ς, the more suitable the connection point is.

2.3. Training Strategy. Based on the pretrained model of VGG-19, we use the prepart of the network of it as the feature extractor, but the parameters of it are fixed, which will not be updated during the training of the hash sequence generation network. Note that, after the testing of equation (6), the second max pooling of VGG-19 is selected as the branch point, and the whole structure of the network and classification branch is shown in Table 1.

First, before the training of VGG-19, the CIFAR-10 training dataset is used in our scheme, and the optimizer is *SGD*. In the training, the training momentum is 0.9, the weight attenuation is 5×10^{-4}, and according to the number of training epochs, the value of the learning rate is constantly adjusted to accelerate the convergence of the network and improve the accuracy of classification.

During the training and testing of the network, the super parameters α_1 and α_2 are both set as 1, at the same time, the optimizer *Adam* is used for the training of the proposed network (II branch), the learning rate is 0.001, and the value of epoch is 200.

The training dataset of the proposed neural network includes η original images, which are randomly selected from the COCO dataset [28]. Each original image would be transformed into 67 similar images through the attack operations given in Table 2, and we also added 67 different images with the original image. The original image, similar images, and different images have been combined into an image group, and therefore, we will obtain η image groups. Each time, we input an image group into proposed network for training.

3. Experimental Results and Comparisons

In order to evaluate the effectiveness and superiority of proposed scheme, the proposed neural network model in the aspects of perceptual robustness, discrimination, performance of content authentication, image classification, and computational complexity was tested and compared. The MSE is used to measure the hash distance as follows:

$$D\left(\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}\right) = \frac{1}{l} \sum_{j=1}^{l} \left| \mathbf{Q}_j^{(1)} - \mathbf{Q}_j^{(2)} \right|^2, \tag{7}$$

where l represents the length of the hash sequence, and $\mathbf{Q}_j^{(1)}$ and $\mathbf{Q}_j^{(2)}$ are j-th values of hash sequences $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$,

Table 1: Structure of the multitask neural network.

Input (32 × 32 RGB image)	
conv3-64 + ReLU	
conv64-64 + ReLU	
MaxPooling	
Conv64-128 + ReLU	
Conv128-128 + ReLU	
MaxPooling	
Branch I	Branch II
Conv128-256 + ReLU	
Conv256-256 + ReLU	Conv128-256 + BN + ReLU
Conv256-256 + ReLU	Conv256-256 + BN + ReLU
Conv256-256 + ReLU	
MaxPooling	MaxPooling
Conv256-512 + ReLU	
Conv512-512 + ReLU	Conv256-512 + BN + ReLU
Conv512-512 + ReLU	Conv512-512 + BN + ReLU
Conv512-512 + ReLU	
MaxPooling	AvgPooling
Conv512-512 + ReLU	
Conv512-512 + ReLU	FC-512 + ReLU
Conv512-512 + ReLU	
Conv512-512 + ReLU	
MaxPooling	
FC-512 + ReLU	
FC-256 + ReLU	FC-50
FC-10	
Softmax	

Table 2: Eight common content retention operations (attack operations) for images.

Operations	Parameter name	Values of parameter
Speckle noise	Variance	0.01, 0.05, 0.1, 0.2, 0.3
Scaling	Scaling ratio	0.2, 0.3, 0.4, 0.5, 1.5, 2, 4
Median filtering	Filter size	1, 3, 5, 7, 9, 11, 13, 15, 17, 19
Circle average filtering	Filter size	1, 5, 10, 15, 20, 25, 30, 35, 40
JPEG compression	Quality factor	1, 5, 10, 30, 50, 70, 90, 95
Rotation and cropping	Rotation angle	1, 2, 3, 4, 5, 6, 8, 10, 12
Gamma correction	Γ	0.55, 0.65, 0.75, 0.85, 0.95, 1.05, 1.15, 1.25, 1.35, 1.45
Gaussian filtering	Variance	0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15, 0.2, 0.25

respectively. During the measurement of hash distance D, if D is smaller than threshold θ, the image pair are defined as similar images; on the contrary, they are defined as different images. Note that, the hardware environment of all experiments was uniform, and the CPU was i9-10900X, the GPU was RTX 2080 Ti, and the RAM was 32 GB.

3.1. Robustness Analysis. In the robustness test, 1,000 images (not in the training dataset) from the COCO dataset [28] were randomly sampled as original images for content retention attack operations. Common robust attacks included speckle noise, median filtering, and rotation and cropping. Specific operations and parameter settings are shown in Table 2. Each image generated 67 perceptually similar images, and a total of 67,000 similar images were obtained. Meanwhile, 67,000 hash distances were obtained by using equation (7). Table 3 shows the extreme value, mean value, and standard deviation of each attack operation. The extreme max and min values are used to indicate the overall numerical range, and the mean and standard deviation values represent the situation of hash distance fluctuations. However, as the robustness testing data of all 1,000 images were difficult to display, five typical standard images, named *Airplane*, *Baboon*, *Boat*, *House,* and *Peppers,* were selected to display, as shown in Figure 2. After the content retention attack operations with eight different parameters as shown in Table 2 were used for these five images, the distance measurement was carried out according to the hash sequence generated from the original image and 67 similar images by using equation (7). Figure 3 shows the changes of 5×67 hash distances and demonstrates the superior robustness of the proposed scheme.

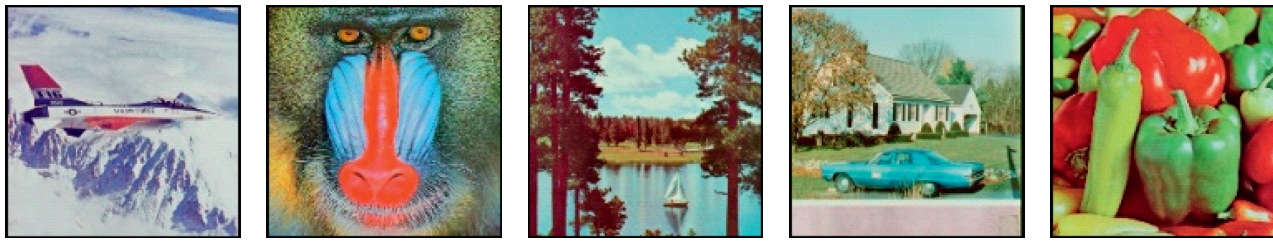
As shown in Figure 3, for JPEG compression, speckle noise, circle average filtering, median filtering, and scaling attack operations, the hash distances are small, less than 0.25. In the Gaussian filtering operation, with the increase of variance, the hash distance of *Boat* increases greatly compared with that of the other four images, but it is still in an acceptable range. Although the rotation and cropping and gamma correction operations have strong changes, compared with other attack operations, the average values of hash distances of these two operations are 0.3724 and 0.2089, respectively. As observed in Table 3, the performance of the proposed scheme was still excellent under rotation and cropping and gamma correction operations.

3.2. Discrimination Capability. To verify the discrimination capability of the proposed scheme, we used the UCID database [29], which contains 1,338 different images with sizes of 512×384 and 384×512. First, we generated hash sequences of the first 1,000 images in [29]. Then, we calculated the hash distance D between each image and other 999 images. So that, we obtained $(1{,}000 \times 999)/2 = 495{,}500$ hash distances. Through the analysis of this experiment, and estimated according to values, the distribution of these hash distances followed the data distribution of mean $\mu = 1.871$ and standard deviation $\sigma = 2.258$. Obviously, the smaller the threshold θ, the lower the collision probability, which means better discrimination capability. When the hash distance D of two images is less than the preset threshold value θ, two images are defined as perceptual similar image pair. If the threshold value θ is too small, the network will distinguish some similar images as different ones, thus affecting the robustness performance of the proposed scheme. Therefore, we need to choose an appropriate threshold θ to achieve a balance between perceptual robustness and discrimination.

As can be seen from Table 3, the average hash distance of perceptually similar images in common image content retention operations is less than 0.4. In addition, when the threshold value $\theta = 0.4$, the collision probability of the

Table 3: Statistics of hash distances under image content retention operations.

Operations	Min.	Max.	Mean	Std.
Speckle noise	1.8×10^{-4}	0.1751	0.0253	0.0377
Scaling	8.0×10^{-5}	0.0550	0.0046	0.0109
Median filtering	0	0.0091	0.0026	0.0025
Circle average filtering	0	0.2417	0.0512	0.0646
JPEG compression	1.3×10^{-4}	0.1138	0.0131	0.0253
Rotation and cropping	1.6×10^{-3}	3.2829	0.3724	0.7473
Gamma correction	5.9×10^{-4}	2.7986	0.2089	0.5312
Gaussian filtering	7.9×10^{-4}	0.1424	0.0268	0.0321



(a) (b) (c) (d) (e)

Figure 2: Five standard images for testing. (a) Airplane. (b) Baboon. (c) Boat. (d) House. (e) Peppers.

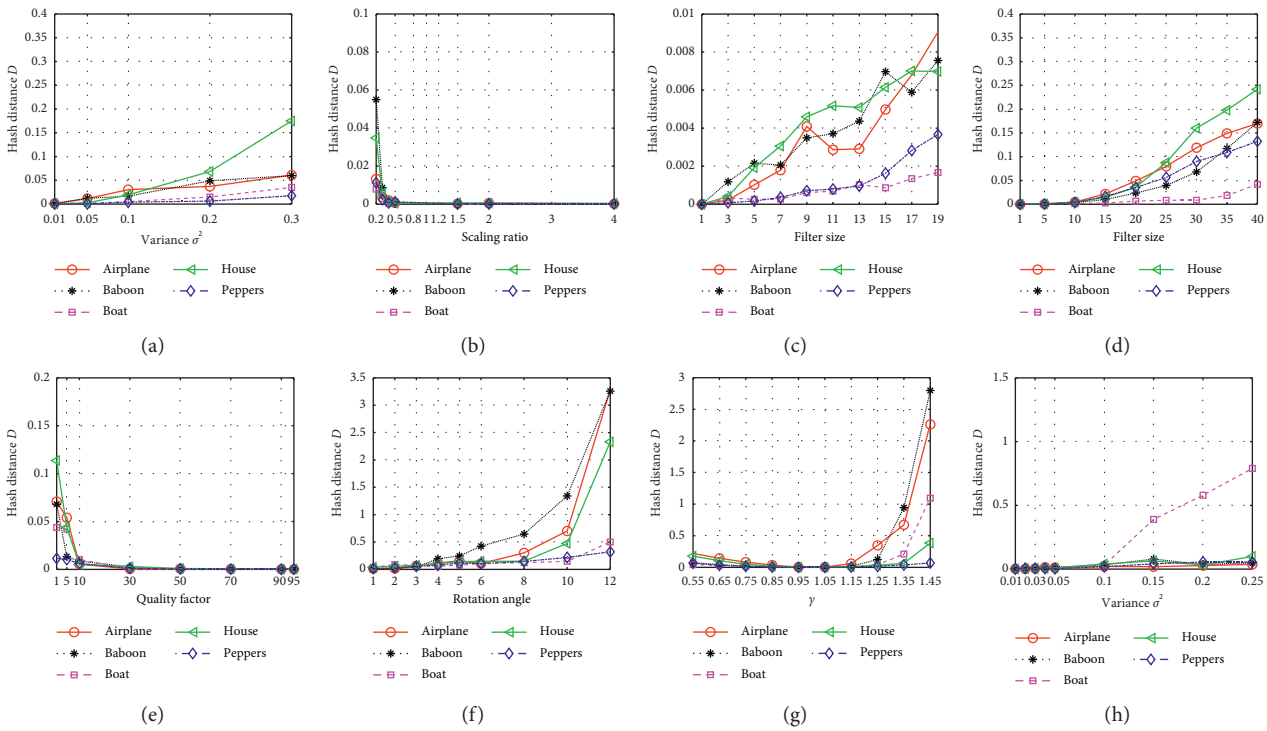


Figure 3: Test curves of five standard images in different image attack operations. (a) Speckle noise. (b) Scaling. (c) Median filtering. (d) Circle average filtering. (e) JPEG compression. (f) Rotation and cropping. (g) Gamma correction. (h) Gaussian filtering.

proposed scheme is 0.0607, which is shown in Table 4; that is, 93.93% of different images are correctly judged. Although hash distances between some perceptually similar images in Table 3 are greater than 0.4, they are only a very small part of the results and have little impact on the overall performance. Therefore, in the proposed scheme, we set the threshold $\theta = 0.4$ to achieve the balance of perceptual robustness and discrimination.

Table 4: The collision probability P_c under different threshold θ

Threshold θ	Collision probability P_c
0.90	0.2741
0.80	0.2284
0.70	0.1827
0.60	0.1385
0.50	0.0972
0.40	0.0607
0.30	0.0308

3.3. Performance of Content Authentication. In order to illustrate the superiority of the proposed scheme, the perceptual robustness and discrimination were considered, and we compared the proposed scheme with other four classical image hashing schemes: DCP [7], RP-IVD [4], RP-NMF [3], and DAE NN-based [24]. The first three schemes used traditional methods, and the last method used deep learning. Because the perceptual robustness and discrimination of the scheme are contradictory, when the perceptual robustness of the scheme is strong, the discrimination must be relatively weak, and vice versa. In order to compare the proposed scheme fairly with the other four schemes, we considered the combined performance of perceptual robustness and discrimination of each scheme as the content authentication ability to perceptually similar and different images.

In this experiment, we randomly selected 1,000 images (not in the training dataset) from [28] to construct a testing image dataset. Each image of this dataset corresponded to 67 perceptual similar images generated by the image content retention operations shown in Table 2 and 67 different images were randomly selected from this dataset. Through setting different values of θ, P_{FAR} and P_{FRR} can be calculated by the following equation:

$$\begin{aligned} P_{\mathrm{FAR}}(\theta) &= \Pr\left(D\left(\mathbf{H}^{(1)}, \mathbf{H}^{(2)}\right) \le \theta\right), \\ P_{\mathrm{FRR}}(\theta) &= \Pr\left(D\left(\mathbf{H}^{(1)}, \mathbf{H}^{(2)}\right) > \theta\right), \end{aligned} \tag{8}$$

where P_{FAR} is the probability of the hash distance less than θ. On the contrary, P_{FRR} is the probability of the hash distance larger than θ, and $\Pr(\cdot)$ is the function of probability. F_1 score is an important quantitative index used to measure the accuracy of content authentication, and when the F_1 score is larger, the performance of content authentication is better. The calculated method of F_1 score is

$$F_1 = \max_{\theta}\left\{\frac{2 \cdot \left[1 - P_{\mathrm{FAR}}(\theta)\right] \cdot \left[1 - P_{\mathrm{FRR}}(\theta)\right]}{\left[1 - P_{\mathrm{FAR}}(\theta)\right] + \left[1 - P_{\mathrm{FRR}}(\theta)\right]}\right\}. \tag{9}$$

Table 5 lists the comparation of F_1 scores of four schemes in different image content retention operations, and bold one of F_1 scores in each line is the best one. The proposed scheme is almost ahead of existing four excellent image hashing schemes in the eight attack operations, and the mean of F_1 score in our scheme is the largest among them, as shown in Table 5. In particular, compared with DAE NN-based [24], four F_1 scores of our scheme are significantly ahead, and other four F_1 scores are also very close. To further demonstrate the overall performance of content authentication based on perceptual robustness and discrimination, we used ROC (receiver operating characteristic) to demonstrate the overall performance of our scheme and other four schemes. In Figure 4, the abscissa is P_{FRR}, the ordinate is $1-P_{\mathrm{FAR}}$, and the ROC curves are closer to top left corner, which means that the better performance of content authentication. Through ROC curves of four schemes in Figure 4, the curve of our scheme is closer to top left corner than other curves. Although the curve of the scheme [24] is close to ours, our scheme is better from the small graph magnified in the lower right. To sum up, according to the quantization results and ROC curves, our scheme had satisfactory performance than [3, 4, 7, 24] in image content authentication.

The generalization ability of our scheme was also tested, different image datasets were used, including COCO dataset [28], UCID dataset [29], Imagenet dataset [30], and NUS_WIDE dataset [31]. ROC curves generated by our scheme for different image datasets are shown in Figure 5. As observed from Figure 5, our scheme has good generalization ability in different image datasets, and at the same time, our scheme has satisfactory adaptive capacity.

3.4. Performance of Image Classification. Because the multitask neural network was used in the proposed scheme, apart from the function of hash authentication, the network also has the function of image classification. We used original images of CIFAR-10 dataset to test the accuracy rate of image classification in our pretrained VGG-19 neural network. There are 1,000 images in this testing dataset. Some results of the image classification of our pretrained network are displayed in Figure 6. The images of the first and second lines are completely correctly identified, but ship (as shown in the red box) is mistakenly identified as truck in the third line, which can be explained that ship had relatively similar characteristics to truck. The accuracy of classification in the final line is 90%. After the test of all images, the final accuracy rate was 93.42%, which means that our multitask neural network can do the task of image classification well.

3.5. Computational Complexity. In actual application, the performance of proposed scheme will face a variety of hardware resources and environment constraints. Therefore, considering the actual use and deployment of the scheme, it is necessary to test and compare the computational complexity of those schemes. In order to avoid the contingency, we randomly selected 100 images from [27] to generate the hash sequences and recorded the running time. Meanwhile, the same test would be carried out on other schemes to compare the advantages. The results are shown in Table 6, although the computational complexity and hash length are not best, considering the influence of multitask function and the tiny gap between ours and the best ones, our scheme is still the leading one in comprehensive performance.

3.6. Application of Tampering Detection. The tampering authentication is also important, and the relative test was also performed for the proposed scheme. Tampered images

Table 5: F_1 scores of different schemes in different content retention operations.

Operations	Proposed	DCP [7]	RP-IVD [4]	RP-NMF [3]	DAE NN-based [24]
Speckle noise	**0.9996**	0.9340	0.9409	0.9299	0.9995
Scaling	**0.9996**	0.8487	0.8909	0.9665	0.9993
Median filtering	0.9996	0.9631	0.9782	0.9891	**0.9998**
Circle average filtering	**0.9988**	0.9063	0.7121	0.7889	0.9895
JPEG compression	0.9993	0.7957	0.9383	0.9925	**0.9998**
Rotation and cropping	**0.9990**	0.6339	0.9936	0.9983	0.9961
Gamma correction	0.9987	0.9518	0.9903	0.8418	**0.9999**
Gaussian filtering	0.9991	0.9195	0.8148	0.8656	**0.9992**
Mean	**0.9992**	0.8708	0.9074	0.9216	0.9979

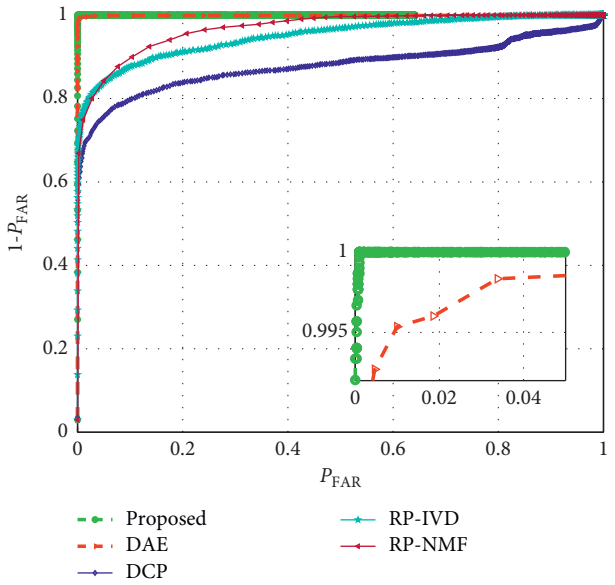


Figure 4: ROC curves about comprehensive performance of five schemes.

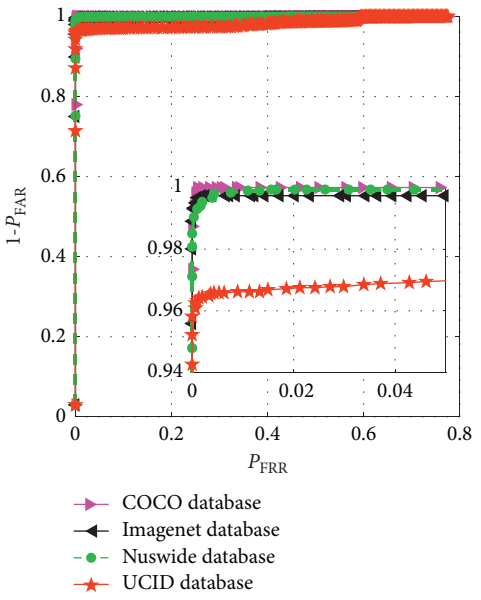


Figure 5: ROC curves of the proposed scheme in each image dataset.

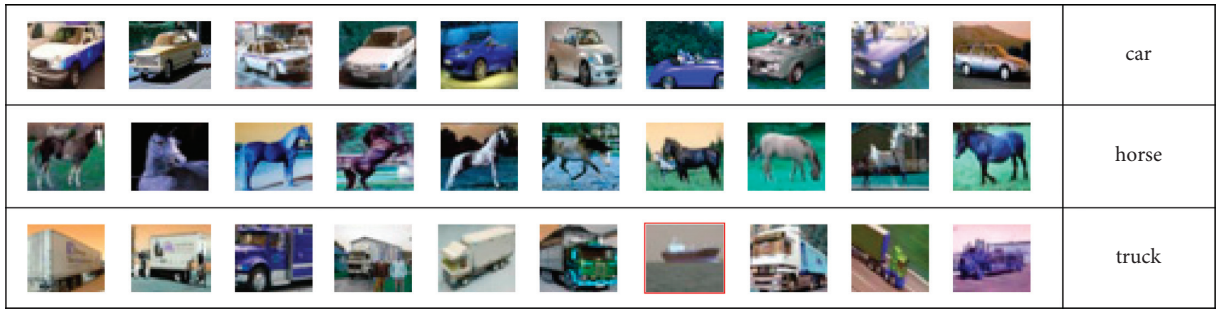


Figure 6: Partial results of image classification.

Table 6: The computational complexity and hash length of each scheme.

Indicator	Proposed	DCP [7]	RP-IVD [4]	RP-NMF [3]	DAE NN-based [24]
Computational complexity (ms)	48.5	154.62	435.2	645.2	17.9
Hash length (digits)	50	64	40	64	50
Multitask function	√	×	×	×	×

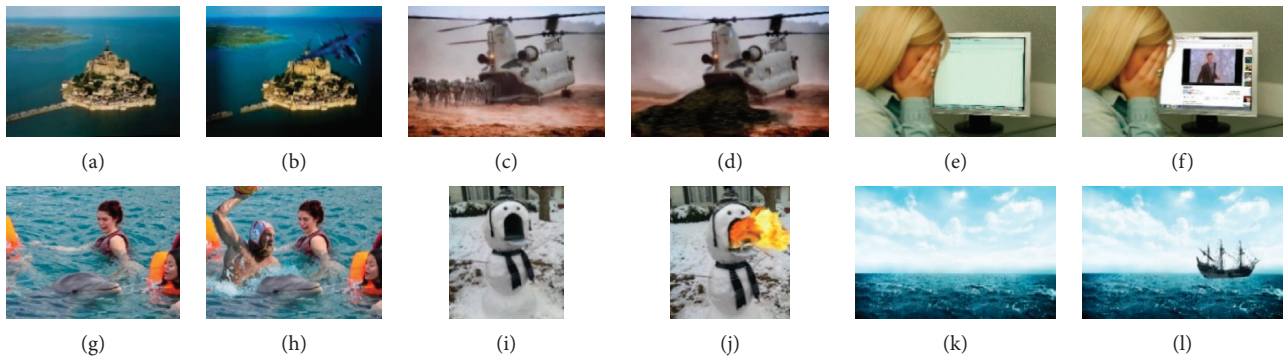


Figure 7: Six pairs of representative tampered test images in the IMD2020 database.

Table 7: Hash distance between the original image and corresponding tampered version.

Image pair	Hash distance D
(a) and (b)	1.36722
(c) and (d)	6.77353
(e) and (f)	1.5763
(g) and (h)	2.50353
(i) and (j)	1.02164
(k) and (l)	1.14891

can be judged by comparing the hash distance between original image and other images efficiently in a large image dataset. In order to test the tampering detection capability of proposed scheme, we randomly selected some original images and corresponding tampered versions from [32]. In Figure 7, we displayed six pair original-tampered images and, in each pair, original image in left, tampered image in right.

The hash distances of different tampered image pairs in Table 7 are larger than $\theta = 0.4$, which indicates the hash sequences are significantly different between the original image and corresponding tampered version. Therefore, the proposed scheme has good tampering authentication performance in practical application, and it can be applied to certain tampering detection scenarios.

4. Conclusion

In order to improve the performance of the image hashing scheme and the reusability of the neural network, we proposed a dual-branch multitask neural network with functions of hash sequence generation and image classification. By looking for the branch point in multiple max pooling layers of VGG-19 network and adding two branch networks, the proposed neural network could have two functions and used one feature extractor. In the two branch networks, one is the original network for image classification after the branch point of the VGG-19 network, and the other is the proposed network to generate the hash sequence. In order to ensure the network converges to the target result, a loss function is proposed to measure the hash distance, which is combined with the MSE and Sigmoid function. The experimental results

show that the proposed scheme has superior robustness and discrimination, and the testing results of image classification are better than the existing classical schemes. The proposed scheme can resist speckle noise, median filtering, rotation and cropping, etc., and it also has advantage in content authentication performance. Through the comparation of ROC curves with other schemes, it can be seen that the proposed scheme is still ahead of them in comprehensive performance. In addition, it has some superiority and applicability in computational complexity and tampering detection applications. In terms of task of image classification, it can be applied to common task of image classification to ensure the function of multitask and improve the applicability of the proposed network to multiple scenarios.

Data Availability

The image datasets used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62172280, U20B2051, and 62172281, in part by the Natural Science Foundation of Shanghai under Grant 21ZR1444600, and in part by the STCSM Capability Construction Project for Shanghai Municipal Universities under Grant 20060502300.

References

[1] M.. Schneider and S.-F. Chang, "A robust content based digital signature for image authentication,"vol. 3, pp. 227–230, in *Proceedings of the 3rd IEEE International Conference on Image Processing*, vol. 3, pp. 227–230, IEEE, Lausanne, Switzerland, sep 1996.

[2] C.-P. Yan, C.-M. Pun, and X.-C. Yuan, "Multi-scale image hashing using adaptive local feature extraction for robust tampering detection," *Signal Processing*, vol. 121, pp. 1–16, 2016.

[3] Z. Tang, X. Zhang, and S. Zhang, "Robust perceptual image hashing based on ring partition and NMF," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 711–724, 2013.

[4] Z. Tang, Z. Huang, X. Zhang, and H. Lao, "Robust image hashing with multidimensional scaling," *Signal Processing*, vol. 137, pp. 240–250, 2017.

[5] Z. Tang, X. Zhang, X. Li, and S. Zhang, "Robust image hashing with ring partition and invariant vector distance," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 200–214, 2015.

[6] X. Nie, X. Li, Y. Chai, C. Cui, X. Xi, and Y. Yin, "Robust image fingerprinting based on feature point relationship mining," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 6, pp. 1509–1523, 2018.

[7] C. Qin, X. Chen, X. Luo, X. Zhang, and X. Sun, "Perceptual image hashing via dual-cross pattern encoding and salient structure detection," *Information Sciences*, vol. 423, pp. 284–302, 2018.

[8] C. Qin, Y. Hu, H. Yao, X. Duan, and L. Gao, "Perceptual image hashing based on weber local binary pattern and color angle representation," *IEEE Access*, vol. 7, no. 45, pp. 460–471, 2019.

[9] Q. Shen and Y. Zhao, "Perceptual hashing for color image based on color opponent component and quadtree structure," *Signal Processing*, vol. 166, 2020.

[10] Z. Huang and S. Liu, "Perceptual hashing with visual content understanding for reduced-reference screen content image quality assessment[J]," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2808–2823, 2020.

[11] X. Wang, M. A. Jianfeng, and Y. Miao, "Efficient privacy-preserving image retrieval scheme over outsourced data with multi-user," *Journal on Communications*, vol. 40, no. 2, pp. 31–39, 2019.

[12] Y. Li, J. Ma, and M. Yinbin, "Encrypted image retrieval in multi-key settings based on edge computing," *Journal on Communications*, vol. 41, no. 4, pp. 14–26, 2020.

[13] J. Fridrich and M. Goljan, "Robust hash functions for digital watermarking," in *Proceedings of the International Conference on Information Technology: Coding and Computing*, pp. 178–183, IEEE, Las Vegas, NV, USA, March 2000.

[14] H. Zhang, H. Zhang, Q. Liu, and X. Niu, "Image perceptual hashing based on human visual system," *Acta Electronica Sinica*, vol. 36, no. 12A, pp. 31–34, 2008.

[15] C. Qin, C.-C. Chang, and P.-L. Tsou, "Robust image hashing using non-uniform sampling in discrete Fourier domain," *Digital Signal Processing*, vol. 23, no. 2, pp. 578–585, 2013.

[16] V. Patil and K. Tanuja, "Image hashing using DWT-CSLBP," *Journal of Computers*, vol. 14, no. 3, pp. 210–222, 2019.

[17] S. Liu and Z. Huang, "Efficient image hashing with geometric invariant vector distance for copy detection," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 4, pp. 1–22, 2019.

[18] C. Qin, X. Chen, D. Ye, J. Wang, and X. Sun, "A novel image hashing scheme with perceptual robustness using block truncation coding," *Information Sciences*, vol. 361-362, pp. 84–99, 2016.

[19] Y.-N. Li and P. Wang, "Robust image hashing based on low-rank and sparse decomposition," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 2154–2158, IEEE, Shanghai, China, March 2016.

[20] Z. Han, M. Long, J. Wang, and Y. Cao, "Deep hashing network for efficient similarity retrieval," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, pp. 2415–2421, AAAI Press, Beijing, China, Feb 2016.

[21] J. Zhang and Y. Peng, "Query-adaptive image retrieval by deep-weighted hashing," *IEEE Transactions on Multimedia*, vol. 20, no. 9, pp. 2400–2414, 2018.

[22] F. Shen, Y. Xu, L. Liu, Y. Yang, Z. Huang, and H. T. Shen, "Unsupervised deep hashing with similarity-adaptive and discrete optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 3034–3044, 2018.

[23] S. Cheng, H. Lai, L. Wang, and J. Qin, "A novel deep hashing method for fast image retrieval," *The Visual Computer*, vol. 35, no. 9, pp. 1255–1266, 2019.

[24] Y. Li, D. Wang, and L. Tang, "Robust and secure image fingerprinting learned by neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 362–375, 2019.

[25] C. Qin, E. Liu, G. Feng, and X. Zhang, "Perceptual image hashing for content authentication based on convolutional neural network with multiple constraints," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 11, pp. 4523–4537, 2021.

[26] B. Chen, W. Tan, G. Coatrieux, Y. Zheng, and Y.-Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment," *IEEE Transactions on Multimedia*, vol. 23, pp. 3506–3517, 2021.

[27] B. Chen, X. Liu, Y. Zheng, G. Zhao, and Y.-Q. Shi, "A robust GAN-generated face detection method based on dual-color spaces and an improved Xception," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, pp. 3506–3517, 2021.

[28] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft coco: common objects in context," *Computer Vision - ECCV 2014*, vol. 8693, pp. 740–755, 2014.

[29] G. Schaefer and M. Stich, "Ucid - an uncompressed color image database," *Proceedings of in Storage and Retrieval Methods and Applications for Multimedia*, pp. 472–480, SPIE, Nottingham, UK, 2004.

[30] D. Jia, W. Dong, R. Socher, Li-J. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the Conference on Computer Vision & Pattern Recognition*, vol. 17, no. 5, pp. 248–255, IEEE, Miami, FL, USA, June 2009.

[31] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "NUS-WIDE: a real-world web image database from National University of Singapore," in *Proceedings of the International Conference on Image & Video Retrieval*, vol. 48, pp. 1–9, ACM, New York, NY, United States, July 2009.

[32] N. Adam, B. Mahdian, and S. Saic, "A large-scale annotated dataset tailored for detecting manipulated images," in *Proceedings of the IEEE Winter Applications of Computer Vision Workshops*, pp. 71–80, IEEE, Snowmass, CO, USA, March 2020.