electronics

MDPI

Article

A Deep Learning Approach in the DCT Domain to Detect the Source of HDR Images

Jiayu Wang [1,2], Hongquan Wang [1], Xinshan Zhu [1,2,*] and Pengwei Zhou [3]

[1] School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China; wangjiayu9697@163.com (J.W.); 1015203049@tju.edu.cn (H.W.)

[2] State Key Laboratory of Digital Publishing Technology, Beijing 100871, China

[3] National Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China; 11932045@zju.edu.cn

* Correspondence: xszhu@tju.edu.cn; Tel.: +86-151-2258-0213

Received: 17 October 2020; Accepted: 29 November 2020; Published: 3 December 2020

Abstract: Although high dynamic range (HDR) is now a common format of digital images, limited work has been done for HDR source forensics. This paper presents a method based on a convolutional neural network (CNN) to detect the source of HDR images, which is built in the discrete cosine transform (DCT) domain. Specifically, the input spatial image is converted into DCT domain with discrete cosine transform. Then, an adaptive multi-scale convolutional (AMSC) layer extracts features related to HDR source forensics from different scales. The features extracted by AMSC are further processed by two convolutional layers with pooling and batch normalization operations. Finally, classification is conducted by a fully connected layer with Softmax function. Experimental results indicate that the proposed DCT-CNN outperforms the state-of-the-art schemes, especially in accuracy, robustness, and adaptability.

Keywords: image forensics; high dynamic range; inverse tone mapping; discrete cosine transform; convolutional neural networks

1. Introduction

With the limitation of bit-depth, the conventional 8-bit digital images cannot accurately reflect the current state of the environment, resulting in a loss of visual information in regions with imprecise exposures [1,2]. To reflect more realistic information, high dynamic range format stores accurate information by using higher bit-depth and floating-point formats [3]. As a consequence, the dynamic range of HDR images can reach 10^4–10^9 orders of magnitude, which far exceeds the dynamic range of low dynamic range(LDR) images [4,5].

With the development of display techniques, some display devices have been able to display HDR contents [6–8]. Meanwhile, HDR images can be easily obtained with the advancement of mobile devices and imaging techniques. Since native HDR sensors have not been widely used, HDR images are mainly obtained from LDR images. There are two common types of HDR images according to the source of HDR images: (1) HDR images synthesized from multiple LDR images of the same scene with different exposures, which are mainly obtained directly through fusion algorithms when shooting images. This type of HDR images are denoted by mHDR [9,10]. (2) HDR images generated by using inverse tone mapping (iTM) to expand the dynamic range of a single LDR image, which are used to replace the existing LDR images [11]. This type of HDR images are denoted by iHDR [12–14]. There is evidence that the mHDR image is indistinguishable from the iHDR image [15–17]. Hence, source forensics of HDR images has become a new problem in the field of image forensics: identifying mHDR images synthesized from multiple exposures and iHDR images generated by iTM from a single LDR image.

Image forensics methods extract features based on numerical values to identify the source of the image or whether the image has been tampered. Identifying the source of an image is an important issue in the field of image forensics. This article is dedicated to solving the problem of HDR image source forensics. The motivation of this paper is to detect the source of HDR images. More specifically, HDR images are mainly divided into mHDR and iHDR according to the source of HDR images. The proposed method is designed to distinguish mHDR images from iHDR images. From the perspective of multimedia security, solving the problem of identifying the source of HDR images can assist in validating the authenticity of the content in images.

Currently, rare research focuses on forensic problems in the HDR domain. All existing HDR source forensic methods are conducted in the spatial domain. According to the way of extracting features, these methods can be divided into two strategies: (1) Manually specified methods extract hand-crafted features and use support vector machine (SVM) to complete classification [18–21]. (2) Convolutional neural network (CNN)-based methods use CNN to automatically extract features related to forensics and determine the type of the input HDR images in an end-to-end way [22]. In this article, a CNN for HDR source forensics is built in the frequency domain, taking advantage of the frequency domain in HDR forensics feature representation. To our best knowledge, this is the first time HDR image source forensics has been conducted in the frequency domain.

The main contribution of this paper is as follows. First, with the aim of using the decorrelation characteristic of DCT to make CNN focus on the features associated with forensics rather than the content of the image, we designed a multi-channel DCT (MC-DCT) module to convert the HDR image in the spatial domain into a DCT coefficients matrix. Second, we construct a multi-scale convolutional layer with different kernel sizes to extract features from different scales, which improves the ability of CNN to extract forensics-related features. Last, the multi-scale features are weighted by a channel attention mechanism, which allows CNN to focus on the channels with more relevant to forensics. Extensive experiments have shown that the performance of the proposed method is significantly improved compared with existing methods.

The remainder of this paper is organized as follows: Section 2 summarizes relevant research on the HDR images source forensics. Section 3 illustrates the architecture of the proposed DCT-CNN in detail. Section 4 describes the details of the datasets used in the experiments and analyses experimental results on different datasets. Section 5 gives the conclusion.

2. Related Works

Only a little literature exists on image forensics in HDR contents. This is because the HDR format is relatively new in the fields of multimedia and signal processing, and the scarcity of HDR image datasets also limits the development of forensics on HDR contents.

As the first work on forensic problems related to HDR contents. Bateman et al. proposed a scheme to extract suitable features for distinguishing tone-mapped HDR images and LDR images using SVM [18]. This work raised a new problem in the field of image forensics: identifying the LDR images obtained from tone-mapped HDR images and the original LDR images. This forensics problem still focused on LDR contents and the scheme proposed was conducted in LDR contents.

Furthermore, Wei et al. proposed a new forensics problem: identifying the mHDR image synthesized from multiple LDR images with different exposures and the iHDR image obtained via inverse tone mapping of a single LDR image [19]. This new problem was related to HDR contents and was named after the problem of HDR source forensics. This work proposed a powerful HDR forensics feature that distinguished mHDR images from iHDR images by using local high-order statistics (LHS) based on fisher scores calculated under the Gaussian mixture model. However, manually specified method cannot fully extract the features related to HDR source forensics. The drawback is that the feature related to HDR source forensics need to be manually designed, which limits the ability of forensics methods to extract features associated with forensics.

With the development of deep learning, more CNN-based methods were applied to image forensics. To overcome the drawback of manually specified methods, Huo et al. used convolutional neural network (CNN) to achieve source forensics of HDR images [22]. In this method, an end-to-end scheme named HDR-CNN was proposed and validated the feasibility of CNN for HDR source forensics. The experimental results showed that by using convolutional neural networks to extract features automatically, the accuracy of HDR source forensics is much better than that of conventional manually specified methods. However, HDR-CNN which is built in the spatial domian tends to extract features related to the content rather than information about forensics, which limits the performance of this method.

In addition to conducting forensics in the spatial domain, some forensics methods were built in the frequency domain to avoid the interference of images content. To use the decorrelation characteristic of DCT, Zhang et al. proposed a CNN-based method of median filtering forensics in the discrete cosine transform domain by converting the images in the spatial domain into data in the frequency domain through DCT [23]. The drawbacks of this method are that some low-frequency and high-frequency DCT coefficients were discarded and the DCT coefficients are given a fixed weight, which limited the performance of this method. Singhal et al. proposed a CNN-based method for detecting manipulation by converting image residuals into DCT domain [24]. The drawbacks of this method are that the DCT is conducted on the Median Filter Residual (MFR) and with no multi-scale module to extract features from different scales. Inspired by these works, we consider developing an effective HDR source forensics method based on CNN in the DCT domain to avoid drawbacks of manually specified methods and the interference of image content in the spatial domain. To avoid the drawbacks of other DCT-based CNNs mentioned above, we introduce a multi-channel discrete cosine transform (MC-DCT) module to keep all the DCT coefficients and the AMSC module to extract multi-scale features.

3. Deep Learning Architecture

Convolutional neural networks can update weights to extract more specific features in the training process. Therefore, the method proposed in this paper is based on CNN to extract features in the DCT domain. For brevity, DCT-CNN is used as the abbreviation for the proposed method.

3.1. Overview of the Proposed CNN Model

Figure 1 shows the basic process of the proposed CNN for identifying the source of HDR images. In the spatial domain, CNNs tend to extract features related to image content, which will interfere with the accuracy of HDR source forensics. The discrete cosine transform has the characteristics of decorrelation, which can make the data structure lose the spatial pixel dependence and reduce the influence of the image content on the accuracy of forensics. Therefore, the digital image in spatial domain needs to be transformed into frequency domain with DCT. In the proposed scheme, multi-channel discrete cosine transform is implemented on every channel of the HDR image to obtain multi-channel DCT coefficients, which are used as input to the network instead of using the pixel values of the image.

First, the input HDR image is first converted to DCT coefficients by multi-channel discrete cosine transform block, and then a convolutional layer named Conv1 extracts features from the DCT coefficient matrix. The extracted features are processed with Batch Normalization (BN) [25] and ReLU as input of adaptive multi-scale convolution module. This part is represented by Frequency Domain Feature Extraction in Figure 1. The adaptive multi-scale feature extraction process is represented by Adaptive Multi-Scale Feature Extraction in Figure 1. The multi-scale features extracted by AMSC module are processed with BN, ReLU, and max pooling. Then, a two-layer convolutional stream with max pooling and activation function is used for high-level feature extraction, represented by Hierarchical Feature Extraction in Figure 1. To introduce the adaptability of input with different sizes to the network, average pooling is used to downsample the feature map to a fixed size. Finally, a fully connected layer with Softmax activation function is used to implement the classification. Table 1

indicates the outline of the proposed DCT-CNN. Multi-channel discrete cosine transform and adaptive multi-scale feature extraction will be discussed in detail in Sections 3.2 and 3.3.

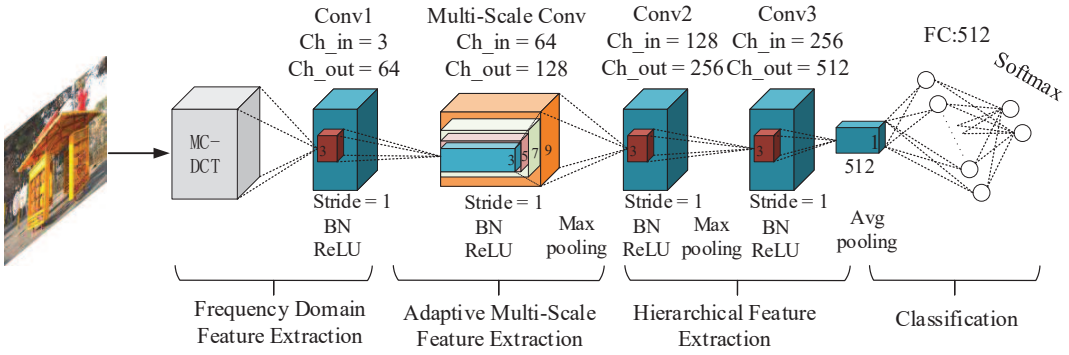


Figure 1. Overview of the proposed CNN in the DCT domain for HDR source forensics.

Table 1. The outline of the proposed network architecture.

Layer	Input	Filter	Stride	Padding	Out
MC-DCT	$32 \times 32 \times 3$	-	-	-	$32 \times 32 \times 3$
Conv1	$32 \times 32 \times 3$	3×3	1	1	$32 \times 32 \times 64$
AMSC	$32 \times 32 \times 3$	3×3 5×5 7×7 9×9	1 1 1 1	1 2 3 4	$32 \times 32 \times 128$
MaxPool	$32 \times 32 \times 128$	2×2	2	0	$16 \times 16 \times 128$
Conv2	$16 \times 16 \times 128$	3×3	1	1	$16 \times 16 \times 256$
MaxPool	$16 \times 16 \times 256$	2×2	2	0	$8 \times 8 \times 256$
Conv3	$8 \times 8 \times 256$	3×3	1	1	$8 \times 8 \times 512$
AvgPool	$8 \times 8 \times 512$	8×8	1	0	$1 \times 1 \times 512$
FC	512×1	-	-	-	2×1

3.2. Multi-Channel Discrete Cosine Transform

The expansion of the dynamic range is mainly carried out on the luminance value of the image. The common operation of the existing HDR source forensics methods is to fuse the red channel (R), the green channel (G) and the blue channel (B) of the HDR image according to Equation (1) to obtain the luminance value of the whole image. Then extract traces related to HDR source forensics based on the distribution of luminance (L).

$$L = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B \tag{1}$$

This approach reduces the dimensionality of input data at the cost of losing part of the information related to HDR source forensics to a certain extent. To improve the accuracy of HDR source forensics, all image information must be fully used. In the proposed method, for the sake of preserving the information in each color channel and converting the input HDR image into the DCT domain, a multi-channel discrete cosine transform as shown in Figure 2 is used. More specifically, the input multi-channel HDR image is split into three channels, denoted by the Red channel, the Green channel and the Blue channel. In addition, DCT is performed on each color channel separately to obtain three individual DCT coefficient matrices. Finally, the three DCT coefficient matrices are concatenated into a 3-channel DCT coefficient matrix. It should be emphasized that the output DCT coefficient matrix has the same size as the input HDR image. Therefore, we can see that the method proposed in this paper is different from the other two DCT-based methods. In Reference [23], the DCT coefficients matrix is multiplied by a weight matrix with values increasing from the upper left corner to the lower right

corner. At the same time, some low-frequency DCT coefficients and high-frequency DCT coefficients in the DCT coefficient matrix are discarded. In Reference [24], the DCT transform is performed on the median filter residual of an image, which discarded the image information before the DCT operation. The method proposed in this paper uses MC-DCT to retain information in multiple color channels without discarding image information of DCT coefficients. Therefore, the proposed method theoretically has better performance than the other two DCT-based methods.

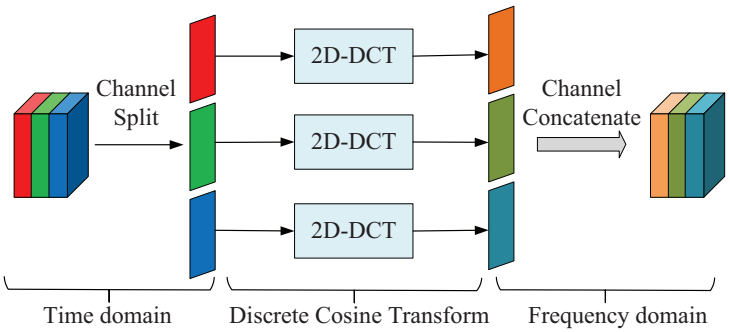


Figure 2. Multi-channel discrete cosine transform.

3.3. Adaptive Multi-Scale Feature Extraction

To represent the forensic features more efficiently, we develop the adaptive multi-scale block, where the convolution operations with n convolution kernels of different sizes are carried out on the input in a parallel manner. Then, multiple scale features are weighted by a channel attention mechanism.

The adaptive multi-scale feature extraction module is shown in Figure 3. Multiple scale features are extracted using a multi-scale convolutional layer, and these features are weighted by a channel attention mechanism. Convolutional layers with different kernel sizes enable CNN to extract features related to HDR source forensics from diverse scales. In this work, channel attention mechanism is used to assign weights to the features extracted by the multi-scale convolutional layer. This adaptive multi-scale feature extraction block can emphasize features that are positive for forensics and suppress irrelevant features by applying channel-wise weights to every channel of multi-scale feature.

Specifically, four convolutional layers with different kernel sizes are carried out on the input to obtain four sets of features that correspond to different scales. Each set of features has 32 channels. Then, a multi-scale feature matrix with 128 channels is derived by concatenating four feature matrices with 32 channels. To extract the most relevant features for HDR source forensics, a channel attention mechanism is used to perform channel-wise weighting operations on the 128-channel multi-scale feature matrix. In this work, the channel attention mechanism is implemented using the Efficient Channel Attention module [26]. In ECA module, global average pooling (GAP) is conducted on the features to obtain aggregated features with a size of $1 \times 1 \times c$, where c denotes the number of channels. Then, a $1D$ convolution is used to extract relationship between channels, followed by a Sigmoid activation to generate the weights of different channels. This channel attention mechanism can be formulated as:

$$w = \sigma(C1D_k(y)) \tag{2}$$

where w refers to the weights of channels, σ is a Sigmoid function, $C1D$ indicates $1D$ convolution, k is the kernel size of convolution. The obtained weights and the input features of the AMSC module are multiplied channel-wise to obtain the weighted multi-scale features. Hence, the subsequent convolutional layers can focus on the channels that are conducive to improving the performance of forensics. The multi-scale structure proposed in [23] involves three different convolution kernel sizes and uses the maxout activation function for activation. During this process, some features related to forensics will be lost. In Reference [24], no multi-scale feature extraction structure is proposed.

Compared with the above DCT-based CNNs, the network proposed in this paper uses an adaptive multi-scale module to extract features without losing features and the features are weighted by a channel attention mechanism to enhance the performance of forensics.

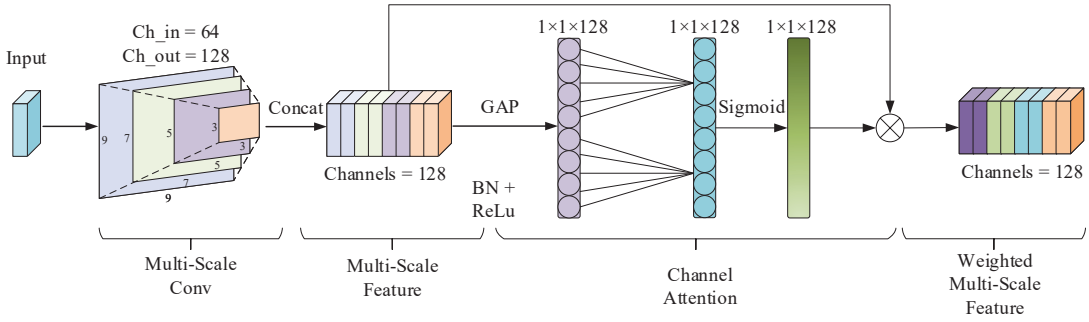


Figure 3. Adaptive multi-scale feature extraction module.

4. Experimental Results

To evaluate the performance of the proposed DCT-CNN on HDR source forensics, we created several datasets with different types of HDR images and different sizes of image blocks. The performance is assessed by classification accuracy (Acc), receiver operating characteristic curve (ROC) and the area under the curve (AUC), and compared with six state-of-art forensics methods in [19–22,24,27]. The classification accuracy (Acc) is defined as:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{3}$$

where TP denotes true positive, which is an outcome when the model correctly predicts the positive class, TN denotes true negative, which is an outcome when the model correctly predicts the negative class, FP denotes false positive, which is an outcome when the model incorrectly predicts the positive class FN denotes false negative, which is an outcome when the model incorrectly predicts the negative class. ROC is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. AUC is defined as the area under the ROC curve enclosed by the coordinate axis.

4.1. Experimental Setup

4.1.1. Training and Testing Datasets

To obtain mHDR images, we choose the following mHDR databases:

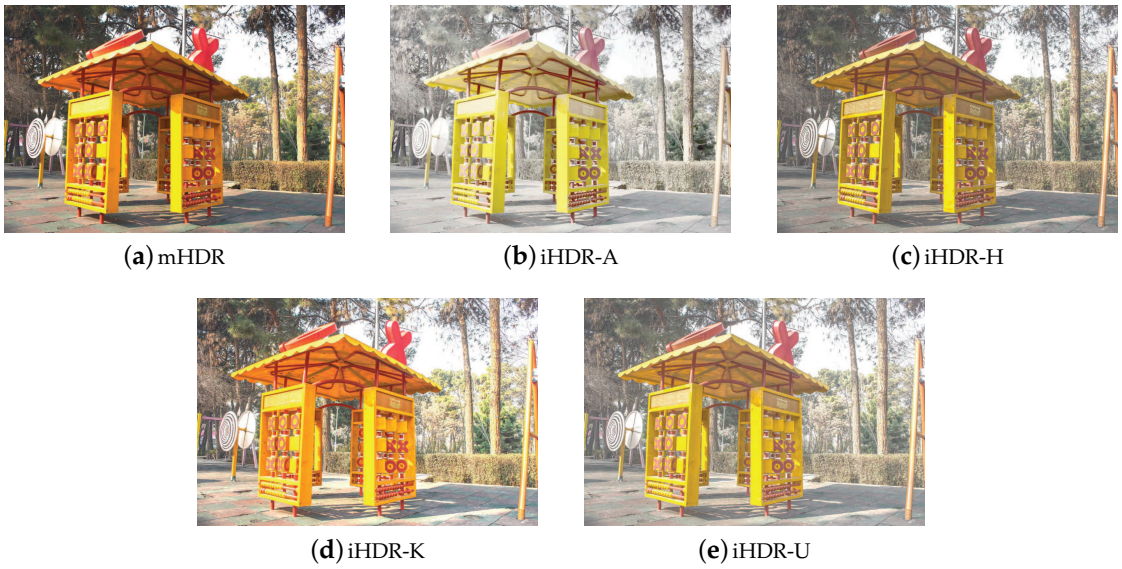
- HDRSID dataset includes 232 mHDR images [28].
- Mantiuk created 8 mHDR images [29].
- Stanford dataset includes 88 mHDR images [30].
- sIBL Archive includes 58 mHDR images [31].

We chose the datasets mentioned above to produce the mHDR image blocks used in the experiments. All HDR images in these datasets were produced using multi-exposure capturing technique. The mHDR images are denoted by 'M'.

The generation of an iHDR image only requires a single LDR image. In this experiment, we chose the MIT-Adobe FiveK dataset [32] as the source of the LDR images. The MIT-Adobe FiveK dataset includes 5000 high-resolution images of different scenes, which can cover a broad range of scenes, subjects, and lighting conditions. In this paper, we select four inverse tone mapping algorithms for generating iHDR images:

1. Akyüz et al.'s method [33], denoted by 'A'. In this method, the input luminance value is first normalized and non-linearly scaled, and then linearly scaled to extend the low dynamic range to the desired high dynamic range.
2. Huo et al.'s method [34], denoted by 'H'. Huo presented a physiological inverse tone mapping algorithm inspired by the property of the Human Visual System (HVS), which could implement the expansion of the dynamic range only in the specific area of the input LDR image. This method can efficiently generate iHDR images with high visual quality.
3. Kovaleski et al.'s method [35], denoted by 'K'. In this work, an inverse tone mapping algorithm based on cross-bilateral filtering was proposed. This method can generate high quality HDR images and videos suitable for a wide range of exposures by using the expand map in specific areas of the image to linearly expand the input LDR content to the desired high dynamic range.
4. Kuo et al.'s method [36], denoted by 'U'. This work proposed an inverse tone mapping method based on histogram. The method includes a content-adaptive inverse tone mapping operator, which has different responses to different scenarios. This algorithm could adaptively select environmental parameters through classification of scenarios to enhance the image in over-exposed areas as well as in remaining well-exposed areas.

We used all 5000 high-quality LDR images from MIT-Adobe FiveK dataset to generate 5000 iHDR images using the above four inverse tone mapping algorithms. As a result, a mHDR dataset including 386 mHDR images and an iHDR dataset including 20,000 iHDR images were obtained. These mHDR and iHDR images constitute the basic experimental datasets. Figure 4 shows the difference between mHDR image and iHDR images generated by different iTM methods.



(**a**) mHDR (**b**) iHDR-A (**c**) iHDR-H

(**d**) iHDR-K (**e**) iHDR-U

Figure 4. Visual comparison of mHDR image with different iHDR images. (**a**) mHDR; (**b**) Akyüz-iHDR; (**c**) Huo-iHDR; (**d**) Kovaleski-iHDR; (**e**) Kuo-iHDR.

Finally, by cropping two type HDR images into blocks of different sizes, specific datasets for evaluating the performance of forensic methods were generated. Specifically, the block size is set to 32, 64, and 128 to verify the performance of forensics under different image sizes. The experiments were conducted on 12 datasets. Each dataset includes 30,000 mHDR image blocks and 30,000 iHDR image blocks. Details of the datasets are shown in Table 2. For each dataset, 25,000 mHDR image blocks and 25,000 iHDR image blocks were randomly selected to form a training set, with the remaining 5000 mHDR images and 5000 iHDR images forming a testing set. After this operation, 12 training datasets and 12 test datasets for subsequent experiments are obtained. These datasets are subsets of the datasets shown in Table 2.

Table 2. Details of the datasets used in experiments.

Size	Dataset	Type	Num	Type	Num
32 × 32	M-A	mHDR	30k	iHDR-A	30k
	M-H	mHDR	30k	iHDR-H	30k
	M-K	mHDR	30k	iHDR-K	30k
	M-U	mHDR	30k	iHDR-U	30k
64 × 64	M-A	mHDR	30k	iHDR-A	30k
	M-H	mHDR	30k	iHDR-H	30k
	M-K	mHDR	30k	iHDR-K	30k
	M-U	mHDR	30k	iHDR-U	30k
128 × 128	M-A	mHDR	30k	iHDR-A	30k
	M-H	mHDR	30k	iHDR-H	30k
	M-K	mHDR	30k	iHDR-K	30k
	M-U	mHDR	30k	iHDR-U	30k

4.1.2. Implementation of the CNN

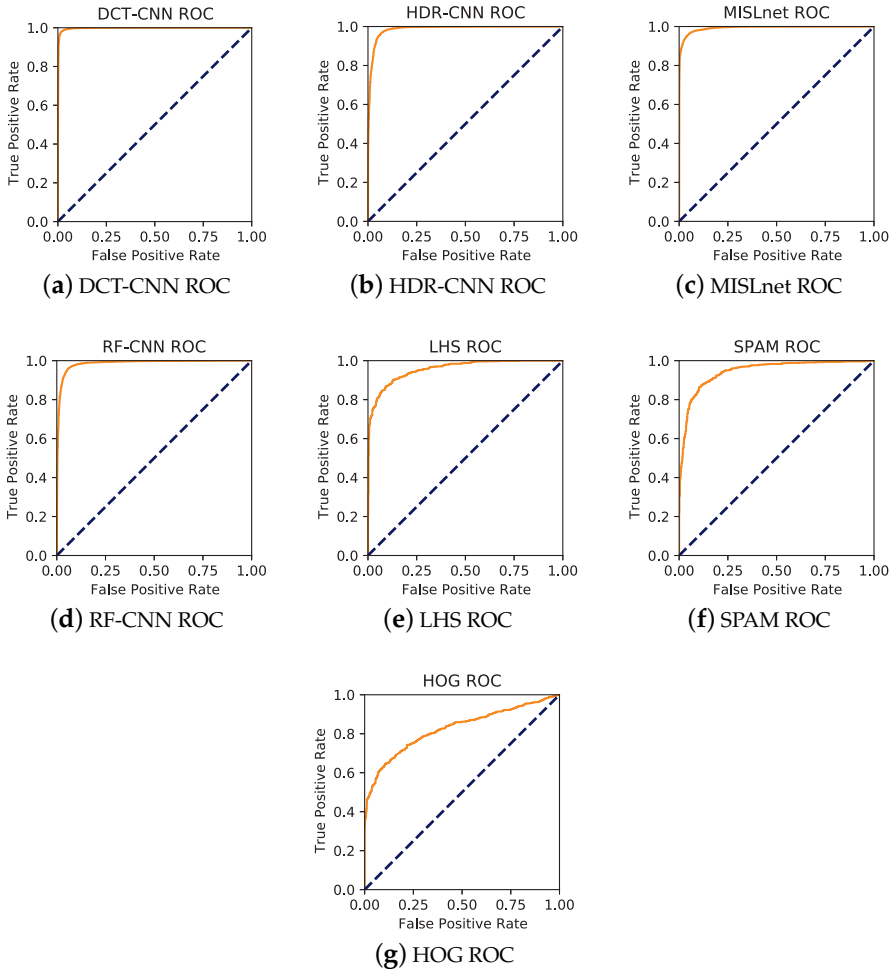
The DCT-CNN for HDR source forensics is implemented with the Pytorch deep learning framework [37]. Experiments were carried out on a high-performance computer with Intel® Core™ i7-9800X (3.80 GHz) (Intel, Santa Clara, CA, USA), 64 GB RAM and NVIDIA® GEFORCE RTX 2080 Ti GPU (NVIDIA, Santa Clara, CA, USA). The parameters of the network are set as follows. The initial learning rate with a learning rate decay strategy is set to 0.001. The batch size is set to 64 images, the loss function is cross-entropy loss, and the optimizer is Adam [38]. Classification accuracy (Acc) is used to evaluate the performance of forensics methods. We chose LHS [19], SPAM [20], HOG [21], HDR-CNN [22], RF-CNN [24] and MISL-net [27] as comparative methods.

4.2. Forensics on Images without Anti-Forensics Attack

The classification accuracy averaged over the test datasets with a resolution of 32 × 32 are summarized in Table 3 for all the tested methods. The best results are marked in bold. Since small-size images include less information related to forensics, experiments conducted on small-size images can reflect the feature extraction capability of forensic methods. Table 3 indicates that the performance of HDR source forensics using manually specified feature extraction methods is weaker than using CNN-based methods to extract features automatically. For instance, the highest classification accuracy of LHS is 88.59% on the M-A dataset, while the accuracy of the two CNN-based forensic methods reached 94.62% and 98.94%. For CNN-based forensic methods, the performance of DCT-CNN in the frequency domain is better than HDR-CNN in the spatial domain. This result validates that the decorrelation of DCT helps CNN extract the most important features related to HDR source forensics. In this experiment, the proposed DCT-CNN manifests the best performance on different HDR datasets. For the proposed DCT-CNN, classification accuracy increased by 10.35% compared with the manually specified feature extraction methods. In addition, compared with HDR-CNN which is a CNN-based forensics method built the spatial domain, the forensics accuracy increased by 4.32%. The experimental results validate that the proposed DCT-CNN for HDR source forensics which is built in the DCT domain can achieve desired forensic performance on 32 × 32 images. It can be observed from Table 3 that compared with other methods, the proposed DCT-CNN gained the highest AUC on different datasets. Figure 5 shows the ROC of different methods, the curve of the DCT-CNN proposed in this paper is closer to the point (0, 1), which indicates that DCT-CNN has better forensics performance over other methods.

Table 3. Forensics accuracy and AUC of different methods on datasets with resolution of 32 × 32.

Image Size	Methods	M-A		M-H		M-K		M-U	
		Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
32 × 32	Proposed	98.94%	0.9969	99.08%	0.9983	99.86%	0.9907	99.53%	0.9974
	HDR-CNN	94.62%	0.9781	90.49%	0.9583	93.57%	0.9654	92.13%	0.9640
	MISLnet	94.04%	0.9779	92.55%	0.9791	91.30%	0.9664	92.03%	0.9683
	RF-CNN	94.75%	0.9702	90.60%	0.9632	90.67%	0.9682	89.93%	0.9637
	SPAM	85.32%	0.9273	83.44%	0.9285	81.37%	0.9165	83.06%	0.9249
	LHS	88.59%	0.9544	87.50%	0.9429	85.47%	0.9485	85.33%	0.9417
	HOG	73.23%	0.8265	75.64%	0.8382	73.17%	0.8249	70.08%	0.8124



(**a**) DCT-CNN ROC (**b**) HDR-CNN ROC (**c**) MISLnet ROC

(**d**) RF-CNN ROC (**e**) LHS ROC (**f**) SPAM ROC

(**g**) HOG ROC

Figure 5. ROC of different methods. (**a**) ROC of the proposed DCT-CNN; (**b**) ROC of HDR-CNN; (**c**) ROC of MISLnet; (**d**) ROC of RF-CNN; (**e**) ROC of LHS; (**f**) ROC of SPAM; (**g**) ROC of HOG.

The classification accuracy and AUC averaged over the test datasets with a resolution of 64 × 64 are summarized in Table 4 for all the tested methods. It can be concluded that in both the CNN-based forensics methods and manually specified feature extraction methods, the accuracy was improved to a certain extent compared with results on 32 × 32 images. Taking LHS as an instance, the forensic accuracy is 93.15% on the M-A dataset with an image size of 64 × 64, while accuracy of LHS on the M-A dataset with an image size of 32 × 32 is 88.59%. The forensics accuracy of HDR-CNN on 64 × 64 images is also improved by 2.92–4.74% compared to result on 32 × 32 images. It should be noted that our proposed method has achieved high forensic accuracy on 32 × 32 images. Hence, performance of proposed DCT-CNN only increased by 0.09–0.49% on 64 × 64 images. In this experiment, the proposed

DCT-CNN still achieves the highest classification accuracy on four different datasets with a resolution of 64 × 64. The DCT-CNN still achieved the highest AUC on different datasets, which verifies its forensic performance from another perspective.

Table 4. Forensics accuracy and AUC of different methods on datasets with resolution of 64 × 64.

Image Size	Methods	M-A		M-H		M-K		M-U	
		Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
64 × 64	Proposed	99.43%	0.9997	99.17%	0.9990	99.54%	0.9946	99.76%	0.9989
	HDR-CNN	97.65%	0.9991	95.41%	0.9783	96.87%	0.9868	96.83%	0.9822
	MISLnet	97.91%	0.9944	96.22%	0.9910	96.61%	0.9906	97.02%	0.9943
	RF-CNN	94.86%	0.9783	91.82%	0.9730	91.66%	0.9795	93.54%	0.9768
	SPAM	87.41%	0.9462	85.24%	0.9342	82.17%	0.9183	85.97%	0.9308
	LHS	93.15%	0.9630	87.93%	0.9828	85.61%	0.9604	84.17%	0.9547
	HOG	79.23%	0.8682	76.64%	0.8651	75.17%	0.8527	72.58%	0.8544

The classification accuracy averaged and AUC over the test datasets with a resolution of 128 × 128 are listed in Table 5. Clearly, 128 × 128 is a relatively large image size. A larger size means that image includes more information related to forensics. It can be observed from Table 5 that the manually specified feature extraction methods and the CNN-based forensics methods have achieved higher classification accuracy on datasets with a resolution of 128 × 128 compared with results on 32 × 32 images and 64 × 64 images. In this experiment, the proposed DCT-CNN still achieves the highest classification accuracy and the highest AUC.

Table 5. Forensics accuracy and AUC of different methods on datasets with resolution of 128 × 128.

Image Size	Methods	M-A		M-H		M-K		M-U	
		Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
128 × 128	Proposed	99.69%	0.9999	99.24%	0.9997	99.72%	0.9978	99.83%	0.9998
	HDR-CNN	98.04%	0.9936	96.14%	0.9854	97.25%	0.9881	97.43%	0.9851
	MISLnet	99.12%	0.9957	98.79%	0.9939	98.82%	0.9975	98.71%	0.9959
	RF-CNN	97.11%	0.9845	95.55%	0.9792	95.32%	0.9874	95.36%	0.9821
	SPAM	90.52%	0.9546	83.19%	0.9417	83.24%	0.9192	86.01%	0.9428
	LHS	93.91%	0.9826	88.29%	0.9873	86.41%	0.9749	86.37%	0.9715
	HOG	81.44%	0.9417	82.73%	0.9485	79.47%	0.9264	81.72%	0.9215

By analyzing the experimental results, we can draw a conclusion that larger image includes more information related to the HDR source forensics. It should be emphasized that among all the methods, RF-CNN and the proposed DCT-CNN were carried out in the DCT domain. The proposed DCT-CNN uses multi-channel DCT to avoid the loss of information and uses an adaptive multi-scale module to extract multi-scale features, which makes the forensic performance of DCT-CNN superior to RF-CNN.

Through Tables 3–5, a conclusion can be drawn that the proposed method is not sensitive to the size of images. High classification accuracy and AUC can also be achieved on the images with low resolution, which validates the strong robustness of DCT-CNN in respect of image size. In addition, we can observe that the performance of forensics methods built in the spatial domain on different types of datasets is not very stable. For instance, HDR-CNN has an accuracy between 90.49–94.62% on different types of datasets with a resolution of 32 × 32. The fluctuation in accuracy of HDR-CNN is 4.13%. The fluctuation in the forensic performance of SPAM, LHS and HOG on datasets with different types is 3.26–5.56%. The fluctuation in the accuracy of our proposed DCT-CNN on datasets with different types are within 1%, which indicates that the proposed DCT-CNN has strong robustness and adaptability in respect of HDR image types.

4.3. Forensics on Images under Anti-Forensics Attack

Image anti-forensics are techniques that aim to make forensics algorithms fail by modifying the images in a visually imperceptible way. Anti-forensics attack are methods used to make forensics method invalid or to decrease the performance of forensics method, which are used to verify the robustness of forensics methods in this experiment. Median filtering has the characteristic of changing the distribution of image pixel values while preserving the content of the image. Due to this characteristic of median filtering, median filtering is often used as anti-forensics attack, which invalidates or reduces the performance of forensic methods. Therefore, it is necessary to study the robustness of the forensics methods under the median filtering attack. The median filter replaces a pixel by the median of all pixels in a neighborhood w:

$$y[m, n] = median\{x[i, j], (i, j) \in w\} \tag{4}$$

where w represents a neighborhood, centered around location $[m, n]$ in the image. Furthermore, in order to verify the robustness against anti-forensics attack of the forensics methods, we chose the median filtering as the anti-forensics attack method.

In this experiment, the size of the images in datasets is fixed to 32×32. Median filtering operation with two different kernels of 3×3 (MF3) and 5×5 (MF5) were conducted on all HDR images. The experiments were conducted on these post-processed datasets to verify the robustness of the HDR source forensics methods. The experimental results are shown in Tables 6 and 7.

Table 6. Forensics accuracy and AUC of different methods on datasets under median flitering (3×3).

Image Size (MF3)	Methods	M-A		M-H		M-K		M-U	
		Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
32×32	Proposed	95.64%	0.9936	95.82%	0.9959	92.45%	0.9706	96.35%	0.9917
	HDR-CNN	92.27%	0.9794	87.39%	0.9726	85.44%	0.9548	87.41%	0.9694
	MISLnet	91.35%	0.9616	89.41%	0.9543	90.15%	0.9528	88.72%	0.9691
	RF-CNN	93.82%	0.9586	87.72%	0.9538	89.57%	0.9417	87.87%	0.9603
	SPAM	76.79%	0.8268	80.31%	0.8243	76.39%	0.8139	76.14%	0.8326
	LHS	80.74%	0.8894	79.06%	0.8816	82.11%	0.8724	73.48%	0.8895
	HOG	69.26%	0.7904	64.89%	0.7858	64.97%	0.7786	71.24%	0.7923

Compared with Table 3, it can be observed from Table 6 that the performance of all forensics methods decreased under the median filtering attack. Especially, the accuracy of forensics methods built in the spatial domain significantly decreased. For instance, LHS gains best performance among manually specified feature extraction methods. However, the accuracy of LHS on the M-A dataset has decreased by 7.85% compared to the accuracy without an attack. As a CNN-based forensics method, HDR-CNN has also decrease by 8.13% on the M-K dataset. For our proposed DCT-CNN, the accuracy under the median filtering attack is still the highest among all methods on four different datasets, which validates that DCT-CNN is robust against anti-forensics attacks.

Table 7. Forensics accuracy and AUC of different methods on datasets under median flitering (5×5).

Image Size (MF5)	Methods	M-A		M-H		M-K		M-U	
		Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
32×32	Proposed	94.21%	0.9859	95.32%	0.9902	90.16%	0.9694	95.38%	0.9896
	HDR-CNN	90.19%	0.9649	84.62%	0.9621	83.33%	0.9607	81.04%	0.9689
	MISLnet	87.73%	0.9503	84.99%	0.9586	84.34%	0.9457	83.80%	0.9621
	RF-CNN	93.15%	0.9014	87.52%	0.9059	88.26%	0.8974	87.22%	0.9146
	SPAM	70.13%	0.7784	73.49%	0.7751	69.72%	0.7546	65.19%	0.7824
	LHS	78.04%	0.8737	78.17%	0.8719	77.35%	0.8637	72.94%	0.8792
	HOG	62.58%	0.7761	62.31%	0.7628	64.55%	0.7549	67.29%	0.7804

Comparing Tables 6 and 7, it can be concluded that the median filtering with a kernel size of 5×5 has a greater impact on the performance of forensics methods than that with a kernel size of 3×3. In the case of more intense anti-forensics attacks, our proposed method still achieved the highest accuracy and the highest AUC on all the datasets. Compared with the results under median filtering with kernel size of 3×3, the accuracy of the forensics methods built in the spatial domain fluctuates between 2–11%, while the fluctuation of DCT-CNN is between 0.5–2.29%, which proves the proposed DCT-CNN is very robust to anti-forensics attacks.

5. Conclusions

In this paper, we propose a CNN-based model in the DCT domain to detect the source of HDR images. To the best of our knowledge, this is the first attempt to achieve HDR source forensics in the frequency domain. Decorrelation of the image content is conducted by transforming the input image in the spatial domain into the DCT domain with a MC-DCT transformation. Hence, the subsequent network can focus on the features related to forensics. Furthermore, an adaptive multi-scale convolution module is applied to extract forensics-related information from different scales with the aim to improve forensics performance of the network. The experimental results show that, compared with the manually specified feature extraction methods and the current CNN-based method, our DCT-CNN has achieved the best classification accuracy and AUC on datasets with different resolutions and datasets with different types of HDR images. Sufficient experiments also validate the strong robustness of the proposed DCT-CNN in respect of image sizes and HDR image types. Moreover, it yields good robustness against median filtering. We hope that this work will inspire follow-up work in the field of HDR source forensics.

Author Contributions: Conceptualization, X.Z.; methodology, X.Z., J.W. and H.W.; software, J.W.; HDR images acquisition, P.Z.; data preprocessing, P.Z.; data analysis, all authors; supervision, X.Z; writing–original draft preparation, J.W. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Grant No. 61972282), and by the Opening Project of State Key Laboratory of Digital Publishing Technology (Grant No. Cndplab-2019-Z001).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this paper:

AMSC	Adaptive Multi-Scale Convolution
Acc	Accuracy
AUC	Area Under the Curve
BN	Batch Normalization
CNN	Convolutional Neural Networks
DCT	Discrete Cosine Transform
HDR	High Dynamic Range
iTM	inverse Tone Mapping
LDR	Low Dynamic Range
MC-DCT	Multi-Channel Discrete Cosine Transform
ROC	Receiver Operating characteristic Curve

References

1. Belyaev, E.; Mantel, C.; Forchhammer, S. Low-complexity compression of high dynamic range infrared images with JPEG compatibility. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
2. Garcia, F.; Schockaert, C.; Mirbach, B. Real-time visualization of low contrast targets from high-dynamic range infrared images based on temporal digital detail enhancement filter. *J. Electron. Imaging* **2015**, *24*, 061103. [CrossRef]

3. Murofushi, T.; Iwahashi, M.; Kiya, H. An integer tone mapping operation for hdr images expressed in floating point data. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2479–2483.
4. Eschbach, R.; Marcu, G.G.; Rizzi, A.; Nezamabadi, M.; Miller, S.; Daly, S.; Atkins, R. Color signal encoding for high dynamic range and wide color gamut based on human perception. *Proc. SPIE Int. Soc. Opt. Eng.* **2014**, *9015*, 90150C.
5. Zhao, H.; Shi, B.; Fernandez-Cull, C.; Yeung, S.; Raskar, R. Unbounded High Dynamic Range Photography Using a Modulo Camera. In Proceedings of the 2015 IEEE International Conference on Computational Photography (ICCP), Houston, TX, USA, 24–26 April 2015; pp. 1–10.
6. Lin, F.C.; Huang, Y.P.; Liao, L.Y.; Liao, C.Y.; Shieh, H.P.D.; Wang, T.M.; Yeh, S.C. Dynamic backlight gamma on high dynamic range LCD TVs. *J. Disp. Technol.* **2008**, *4*, 139–146.
7. Chen, H.; Zhu, R.; Li, M.C.; Lee, S.L.; Wu, S.T. Pixel-by-pixel local dimming for high-dynamic-range liquid crystal displays. *Opt. Express* **2017**, *25*, 1973–1984. [CrossRef] [PubMed]
8. Tan, G.; Huang, Y.; Li, M.C.; Lee, S.L.; Wu, S.T. High dynamic range liquid crystal displays with a mini-LED backlight. *Opt. Express* **2018**, *26*, 16572–16584. [CrossRef] [PubMed]
9. Debevec, P.E.; Malik, J. Recovering high dynamic range radiance maps from photographs. In Proceedings of the ACM SIGGRAPH 2008 Classes, Los Angeles, CA, USA, 11–15 August 2008; pp. 1–10.
10. Tomaszewska, A.; Mantiuk, R. Image Registration for Multi-Exposure High Dynamic Range Image Acquisition. 2007. Available online: http://core.ac.uk/download/pdf/295558346.pdf (accessed on 1 December 2020).
11. Banterle, F.; Ledda, P.; Debattista, K.; Chalmers, A. Inverse tone mapping. In Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, Kuala Lumpur, Malaysia, 29 November–2 December 2006; pp. 349–356.
12. Meylan, L. Tone Mapping for High Dynamic Range Images. 2006. Available online: http://infoscience.epfl.ch/record/86005 (accessed on 1 December 2020).
13. Kinoshita, Y.; Shiota, S.; Kiya, H. Fast inverse tone mapping with Reinhard's global operator. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 1972–1976.
14. Lee, S.; Hwan An, G.; Kang, S.J. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 596–611.
15. Ning, S.; Xu, H.; Song, L.; Xie, R.; Zhang, W. Learning an inverse tone mapping network with a generative adversarial regularizer. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 1383–1387.
16. Kim, S.Y.; Oh, J.; Kim, M. Deep sr-itm: Joint learning of super-resolution and inverse tone-mapping for 4k uhd hdr applications. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 3116–3125.
17. Rempel, A.G.; Trentacoste, M.; Seetzen, H.; Young, H.D.; Heidrich, W.; Whitehead, L.; Ward, G. Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph. TOG* **2007**, *26*, 39. [CrossRef]
18. Bateman, P.J.; Ho, A.T.; Briffa, J.A. Image forensics of high dynamic range imaging. In Proceedings of the 2011 International Workshop on Digital Watermarking, Atlantic City, NY, USA, 23–26 October 2011; pp. 336–348.
19. Fan, W.; Valenzise, G.; Banterle, F.; Dufaux, F. Fine-grained detection of inverse tone mapping in HDR images. *Signal Process.* **2018**, *152*, 178–188. [CrossRef]
20. Pevny, T.; Bas, P.; Fridrich, J. Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 215–224. [CrossRef]
21. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
22. Huo, Y.; Zhu, X. High dynamic range image forensics using cnn. *arXiv* **2019**, arXiv:1902.10938.
23. Zhang, J.; Liao, Y.; Zhu, X.; Wang, H.; Ding, J. A Deep Learning Approach in the Discrete Cosine Transform Domain to Median Filtering Forensics. *IEEE Signal Process. Lett.* **2020**, *27*, 276–280. [CrossRef]

24. Singhal, D.; Gupta, A.; Tripathi, A.; Kothari, R. CNN-based Multiple Manipulation Detector Using Frequency Domain Features of Image Residuals. *ACM Trans. Intell. Syst. Technol. TIST* **2020**, *11*, 1–26. [CrossRef]
25. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
26. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
27. Bayar, B.; Stamm, M.C. Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2691–2706. [CrossRef]
28. High Dynamic Range Image Dataset. Available online: http://faculties.sbu.ac.ir/~moghaddam/index.php/main/page/10 (accessed on 1 December 2020).
29. HDR Image Gallery. Available online: http://pfstools.sourceforge.net/hdr_gallery.html (accessed on 1 December 2020).
30. Xiao, F.; DiCarlo, J.M.; Catrysse, P.B.; Wandell, B.A. High dynamic range imaging of natural scenes. In Proceedings of the Color and Imaging Conference. Society for Imaging Science and Technology, Scottsdale, AR, USA, 12 November 2002; Volume 2002, pp. 337–342.
31. Free HDRI Sets for Smart Image-Based Lighting. Available online: http://www.hdrlabs.com/sibl/archive.html (accessed on 1 December 2020).
32. Bychkovsky, V.; Paris, S.; Chan, E.; Durand, F. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. In Proceedings of the Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 20–25 June 2011.
33. Akyüz, A.O.; Fleming, R.; Riecke, B.E.; Reinhard, E.; Bülthoff, H.H. Do HDR displays support LDR content? A psychophysical evaluation. *ACM Trans. Graph. TOG* **2007**, *26*, 38-es. [CrossRef]
34. Huo, Y.; Yang, F.; Dong, L.; Brost, V. Physiological inverse tone mapping based on retina response. *Vis. Comput.* **2014**, *30*, 507–517. [CrossRef]
35. Kovaleski, R.P.; Oliveira, M.M. High-quality reverse tone mapping for a wide range of exposures. In Proceedings of the 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, Rio de Janeiro, Brazil, 26–30 August 2014; pp. 49–56.
36. Kuo, P.H.; Tang, C.S.; Chien, S.Y. Content-adaptive inverse tone mapping. In Proceedings of the 2012 Visual Communications and Image Processing, San Diego, CA, USA, 27–30 November 2012; pp. 1–6.
37. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv* **2019**, arXiv:1912.01703; pp. 8026–8037.
38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.