DOI:10.16136/j.joel.2023.05.0484

# 基于密集连接和特征消冗网络的零水印方法

何灵强[1],骆 挺[1,2*],李 黎[3],何周燕[1,2],徐海勇[1]

(1.宁波大学 信息科学与工程学院,浙江 宁波 315211;2.宁波大学 科学技术学院,浙江 宁波 315000;3.杭州电子科技大学,浙江 杭州 310000)

**摘要**:针对鲁棒水印不可见性和鲁棒性的矛盾,提出了一种基于密集连接和特征消冗网络(dense connection and redundant feature elimination network,DCRFENet)的零水印方法。首先,为了抵抗不同图像攻击,设计密集连接模块,即从不同卷积层提取浅层和深层图像的鲁棒特征。同时,为了增强零水印的唯一性,结合特征间权重学习与特征内权重学习设计特征消冗模块,从而消除冗余特征以及增强图像的有效特征。其次,融合有效特征与鲁棒特征,生成图像特征图,并进行抗攻击训练。最后,基于训练的DCRFENet,将特征图进行分块,比较分块均值与块内每一特征值的大小构造零水印。实验结果表明,在CIFAR10、COCO、VOC数据集上抵抗单一攻击的平均比特误差率(bit error rate, *BER*)均低于0.03。此外,与现有方法相比,提出的零水印方法对训练的攻击、非训练的攻击以及混合攻击均具有较好的鲁棒性

**关键词**:零水印;唯一性;鲁棒性;密集连接;特征消冗

**中图分类号**:TP309.7 **文献标识码**:A **文章编号**:1005-0086(2023)05-0543-11

# A zero-watermarking method based on dense connection and redundant feature elimination network

HE Lingqiang[1], LUO Ting[1,2*], LI Li[3], HE Zhouyan[1,2], XU Haiyong[1]

(1. Faculty of Information Science and Engineering, Ningbo University, Ningbo, Zhejiang 315211, China; 2. College of Science and Technology, Ningbo University, Ningbo, Zhejiang 315000, China; 3. Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China)

**Abstract**: To address the contradiction between watermarking invisibility and robustness, a zero-watermarking method based on dense connection and redundant feature elimination network (DCRFENet) is proposed. Firstly, in order to resist different image attacks, the dense connection module is designed to extract shallow and deep layer image robust features from different convolution layers. Meanwhile, to enhance the uniqueness of zero-watermarking, the redundant feature elimination module is presented to eliminate redundant information and enhance image valid features by learning weight of inter-feature and intra-feature. Secondly, valid features and robust features are fused to generate the final image feature map, which is used for robustness training. Finally, based on training of DCRFENet, the image feature map is divided into blocks, and zero watermark is obtained by comparing each feature value within the block with its average value. The experimental results show that the average bit error rate (*BER*) is lower than 0.03 for CIFAR10, COCO and VOC datasets. Moreover, compared with the existing methods, the proposed zero-watermarking model is robuster to trained attacks, untrained attacks and hybrid attacks.

**Key words**: zero-watermarking; uniqueness; robustness; dense connection; redundant feature elimination

* **E-mail**: luoting@nbu.edu.cn

**收稿日期**:2022-06-29 **修订日期**:2022-10-20

**基金项目**:国家自然科学基金(61971247,62171243,61501270)、浙江省自然科学基金(LY22F020020,LQ23F010011)、宁波市自然科学基金(2021J134,2022J136,2022J066)和浙江省教育厅科研科研项目(Y202248989)资助项目

## 0 引　言

随着计算机技术和互联网技术的发展，图像、音频、视频等数字媒体能够轻易地被获取。由于数字媒体具有可移植性和易传输性的特点，因此部分非法用户在未获得拥有者许可的情况下，可能对数字作品进行非法篡改、复制和传播，将严重侵犯作品所有者的权利[1]。如何有效地保护图像版权是信息安全领域的一项重要挑战，水印技术为解决图像安全问题的有效方法之一[2]。

水印技术一般分为脆弱水印和鲁棒水印，分别用于图像认证与图像的版权保护[3]。由于鲁棒水印能够抵抗不同图像的攻击，因此常常被用于图像的版权保护。为了提高水印的鲁棒性，一般会基于变换域和矩构建鲁棒水印。然而，鲁棒水印由于额外信息的嵌入，将降低图像的质量。此外，在军事、医疗以及法政等特殊领域，不允许对图像进行任何修改，因此，研究者们提出了零水印技术，该技术在不破坏图像任何质量的情况下，能够对图像进行有效的版权保护，也较好解决了水印不可见性和鲁棒性的矛盾[4-6]。

零水印在图像中提取相应的鲁棒特征构造唯一标识图像的相关信息，从而保护图像的版权。熊祥光等人将图像进行分块，利用分块整体均值与分块均值间的大小关系构建零水印，该方法能够较好抵抗一般图像处理的攻击，但是无法抵抗几何攻击，如旋转攻击和剪切攻击等[7]。此外，采用矩的方式能够较好提取鲁棒以及可区分的图像特征，从而提高零水印的抗攻击能力以及唯一性[8,9]。然而，基于空域特征、变换域特征以及矩特征的零水印方法并不具有抵抗不同图像攻击的普适性，对混合攻击的抵抗能力也较弱。

近些年，由于深度学习能够提取图像的关键特征，因此在计算机视觉领域得到了广泛的应用[10]。除了应用于计算机视觉以外，研究者也将深度学习与水印相结合，从而提高水印性能。然而，这些水印网络模型仍旧不能较好解决水印不可见性与鲁棒性的矛盾，即当鲁棒性提高的同时将较大程度地降低图像质量。

利用深度学习设计零水印是提高图像攻击普适性以及不破坏图像质量的有效方法。EHSAEE 等人提出基于学习轮廓检测的零水印方法，与传统方法（基于 Canny 边缘检测和形态学膨胀）相比，具有较好的鲁棒性[11]。HAN 等[12]利用预训练的 VGG19（visual geometry group network）与离散傅里叶变换（discrete Fourier transform，DFT）构建鲁棒特征，然后采用感知哈希方法（perceptual hashing algorithm，PHA）生成零水印，该方法对局部的非线性几何攻击具有较好的鲁棒性。ATOANY 等人提出端到端的卷积神经零水印网络，能够有效抵抗 JPEG 压缩、滤波等图像处理，但无法有效抵抗旋转等几何攻击，且对组合攻击的鲁棒性较差[13]。

由于密集连接网络能有效利用特征，加强特征传递，缓解网络梯度消失及增强特征鲁棒性等多种优势，因此本文提出了基于密集连接和特征消冗网络（dense connection and redundant feature elimination network，DCRFENet）的零水印方法。首先，利用密集连接将各网络层输出的鲁棒特征进行融合，提高零水印抵抗不同攻击的普适性。其次，为了提高不同图像构造零水印的唯一性，设计特征消冗模块，通过特征间权重学习和特征内权重学习消除特征的冗余信息，增强特征的有效性以及可区分性。最后，为了提高零水印鲁棒性，生成噪声图像进行抗攻击训练，并利用 DCRFENet 获得的最终特征图构造零水印。实验结果证明基于 DCRFENet 的零水印具有较好的唯一性，相比现有零水印方法，具有更强的鲁棒性。本文的主要贡献包括：

1）为了提高零水印的鲁棒性，DCRFENet 的密集连接模块能够从不同卷积层提取不同的浅层和深层的图像特征，作为零水印构造的鲁棒特征。

2）为了增强零水印的唯一性，DCRFENet 的特征消冗模块采用特征间以及特征内权重学习，降低了图像特征的冗余信息，提高不同图像特征的区分度。

3）实验结果表明 DCRFENet 对未经训练的攻击及混合攻击均表现出较强的鲁棒性。相比现有传统零水印方法以及深度学习的方法，基于 DCRFENet 的零水印具有更好的抗攻击泛化能力。

## 1　提出的零水印方法

为了获得零水印的唯一性与鲁棒性，提出基于 DCRFENet 的零水印方法。将大小为 $M\times N\times 3$ 的原始图像与噪声图像分别输入 DCRFENet，提取相应的图像特征，并通过特征损失函数（$L_{\mathrm{MSE}}$）训练对不同图像攻击的抵抗能力，如图 1 所示。基于训练后的 DCRFENet，提取图像特征图，构造相应的鲁棒零水印；同理，受到攻击的图像，利用 DCRFENet 计算图像特征图并提取零水印，验证图像版权。

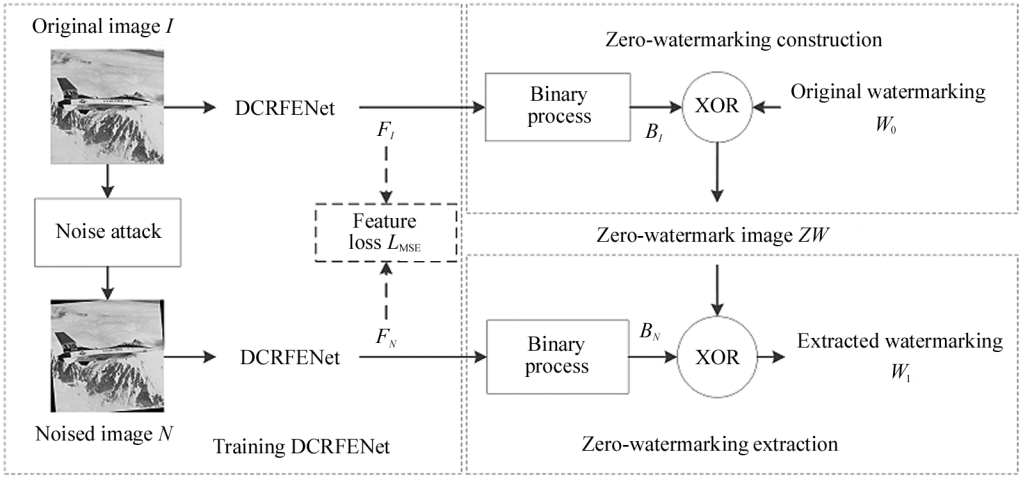


**图 1　基于密集连接和特征消冗网络的零水印框架**

**Fig. 1　Zero-watermarking framework based on dense connection and redundant feature elimination network**

### 1.1　DCRFENet 网络结构

DCRFENet 的目的是为了提取图像的有效特征,并能够抵抗不同的图像噪声攻击,该网络主要包括密集连接模块和特征消冗模块,如图 2 所示。密集连接模块为通过卷积网络生成图像的鲁棒特征,从而保证构造的零水印具有较强的鲁棒性;而特征消冗模块则为了降低冗余特征,增加不同图像特征的可区分性,从而增强零水印的唯一性,即不同图像生成不同的零水印。

#### 1.1.1　密集连接模块

HUANG 等[14]提出的密集卷积网络具有特征传播功能,能够融合各个网络层的特征,从而获得图像的主要能量。受此启发,设计了基于密集连接的网络模块,用于提取图像的鲁棒特征。具体为,密集连接模块中每一个网络层的输出特征图都将作为后面网络层的输入,从而使网络中各层之间达到最大信息流,加强了各个网络层的特征传播。通过 $l$ 层的特征传播,获得图像特征图 $F_d$,$F_d$ 融合了不同深浅网络层的图像信息。

密集连接模块包含 4 个网络层,并且各层之间进行密集连接,保证了特征的重复利用。具体为,设 $F_0$ 为密集连接模块输入的特征图,经过第 1 层,网络的输出特征图为:

$$F_1 = G(F_0)\ , \tag{1}$$

式中,$G(\cdot)$表示网络层的卷积块操作,包括 3×3 卷积函数 $Conv(\cdot)$、批量归一化(Batch Normalization,BN)以及指数线性单元(Exponential Linear Units,ELU)激活函数。$Conv(\cdot)$具体计算为:

$$u_{i,j}^{s} = \sum_{r=1}^{Cl}\sum_{n=-1}^{1}\sum_{m=-1}^{1} u_{i,j}^{r} g_{n+1,m+1}^{r} + b_s\ , \tag{2}$$

式中,$u_{i,j}^{r}$ 为输入特征图中第 $r$ 个通道上位置为$(i,j)$的值,$u_{i,j}^{s}$ 为输出特征图中第 $s$ 个通道位置为$(i,j)$的值,$b_s$ 为其在第 $s$ 个通道上的偏置值,$g_{n+1,m+1}^{r}$ 为卷积核在第 $r$ 个通道位置为$(n+1,m+1)$的权重。

同理,经过第2、3、4层,网络的输出特征图分别为:

$$\begin{cases} F_2 = G(con(F_0,F_1)) \\ F_3 = G(con(F_0,F_1,F_2)) \\ F_4 = G(con(F_0,F_1,F_2,F_3)) \end{cases}\ , \tag{3}$$

式中,$con(\cdot)$表示特征的连接。设每个网络层经过 $G(\cdot)$处理后生成的特征为 $k$ 个,则第 4 层的输入特征有 $k\times(4-1)+k_0$ 个,其中,$k_0$ 为 $F_0$ 的输入特征个数。为了避免输入特征数过大而导致模型过于复杂,每层 32 个卷积核,最终,将特征图 $F_0$、$F_1$、$F_2$、$F_3$ 以及 $F_4$ 进行连接,得到大小为 $H\times W\times C$ 的特征图 $F_d$,其中 $H=M/2$,$W=N/2$,$C=160$,即共有 160 个特征。

#### 1.1.2　特征消冗模块

尽管在密集连接模块,4 层网络获取的特征图 $F_d$ 融合了各个网络层的鲁棒特征。然而,$F_d$ 也包含了图像较多的冗余信息,这些冗余信息会降低图像的可区分性,从而影响零水印的唯一性。因此,提出了特征消冗模块,该模块通过特征间权重学习和特征内权重学习,降低相应冗余信息,从而减少图像的冗余特征,增强有效特征。其中,特征间权重学习包含特征间压缩、共享多层感知器(multilayer perceptron,MLP)和特征间加权融合,如图2所示。特征

压缩目的为提取各个特征的全局信息从而进行 MLP 学习，分别使用全局最大池化层和全局均值池化层，将特征图 $F_d$ 在空间上分别压缩为 $1\times1\times C$ 大小的特征图：

$$\begin{cases} F^c_{\max} = MaxPool(F_d) \\ F^c_{\text{avg}} = AvgPool(F_d) \end{cases}, \tag{4}$$

式中，$MaxPool(\cdot)$ 和 $AvgPool(\cdot)$ 分别表示全局最大池化和全局均值池化函数。MLP 用于学习 $F_d$ 中各个特征之间的权重，并通过权重的大小来衡量特征间相关性。MLP 包含两层全连接网络：第一层的神经元个数为 $C/R$，其中 $R$ 为压缩率，并采用修正线性单元(rectified linear units, ReLu)作为激活函数；第二层神经元个数为 $C$。$F^s_{\max}$ 和 $F^c_{\text{avg}}$ 经过两层全连接神经网络，并依次通过加法融合、Sigmoid 激活函数提取各个特征的权重。相应的具体计算为：

$$\begin{cases} y^1_j = f_{\text{ReLu}}(x^1_j) = f_{\text{ReLu}}\left(b^1_j + \sum_{i=1}^{C}(a^0_{ij}y^0_i)\right) \\ y^2_m = x^2_m = b^2_m + \sum_{j=1}^{C/R}(a^1_{jm}y^1_j) \end{cases}, \tag{5}$$

式中，$y^0_i$ 为 $F^c_{\max}$ 的第 $i$ 个通道值，$x^1_j$ 和 $y^1_j$ 分别是 MLP 第一层中第 $j$ 个神经元的输入值和输出值，$b^1_j$ 为相应的偏置值，$y^2_m$ 为 MLP 第二层中第 $m$ 个神经元的输出值，$f_{\text{ReLu}}(\cdot)$ 为 ReLu 激活函数，$a^1_{jm}$ 为第一层中第 $j$ 个神经元与第二层中第 $m$ 个神经元之间的权重。同理，$F^s_{\text{avg}}$ 经过 MLP 输出的第 $m$ 个神经元输出值为 $z^2_m$。最后，通过 Sigmoid 激活函数将 $y^2_m$ 和 $z^2_m$ 进行融合，得到特征图 $F_d$ 的第 $m$ 个特征权重 $w_m$：

$$w_m = S(y^2_m + z^2_m), \tag{6}$$

式中，$S(\cdot)$ 表示 Sigmoid 激活函数，通过计算各个特征权重 $w_m$，最后形成特征权重图 $M_c(F_d)$。

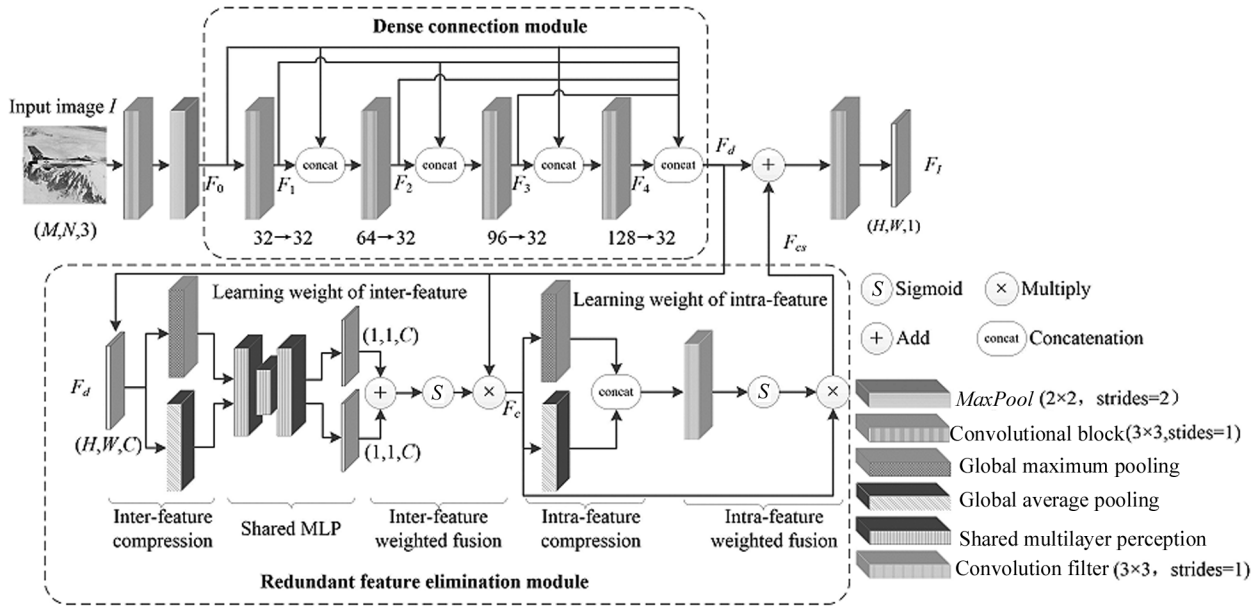


**图 2　密集连接和特征消冗网络(DCRFENet)**

**Fig. 2　Dense connection and redundant feature elimination network (DCRFENet)**

特征加权融合将密集连接模块的输出特征图 $F_d$ 与其特征间的权重 $M_c(F_d)$ 进行加权，降低了特征间冗余信息。具体为将特征图 $F_d$ 中第 $m$ 个通道中位置为 $(i, j)$ 的特征值 $u^m_{i,j}$ 更新为：

$$v^m_{i,j} = u^m_{i,j} \times w_m, \tag{7}$$

式中，$v^m_{i,j}$ 为更新的特征值。同理，相应的特征图 $F_d$ 更新为 $F_c$：

$$F_c = F_d \times M_c(F_d)。 \tag{8}$$

特征内权重学习则为了降低特征内的冗余信息，其包括特征内压缩以及特征内加权融合。首先，特征内压缩分别采用全局最大池化和全局均值池化在通道域上对 $F_c$ 进行压缩，压缩后的特征图大小均为 $H\times W\times1$。特征内加权融合将压缩后的特征图依次通过卷积与 Sigmoid 激活函数得到特征内不同位置的权重：

$$M_s(F_c) = S(Conv(con(F^{cs}_{\max}, F^{cs}_{\text{avg}}))), \tag{9}$$

式中,$F_{\max}^{cs}$和$F_{\text{avg}}^{cs}$分别为:

$$\begin{cases} F_{\max}^{cs} = MaxPool(F_c) \\ F_{\text{avg}}^{cs} = AvgPool(F_c) \end{cases} 。 \tag{10}$$

将 $F_c$ 与 $M_s(F_c)$加权得到特征图 $F_{cs}$:

$$F_{cs} = F_c \times M_s(F_c) 。 \tag{11}$$

$F_{cs}$是在原有特征图 $F_d$ 基础上,改变了原有的特征分布,消除了特征间及特征内的冗余,较大程度增强图像的有效特征,从而保证图像特征的唯一性。

此外,为了保留原有 $F_d$ 的鲁棒性,最终得到的特征图 $F_I$ 为:

$$F_I = Conv(F_{cs} + F_d) 。 \tag{12}$$

特征间权重学习降低了各特征之间的冗余信息,而特征内权重学习减弱了特征内各个区域的冗余信息。相比 $F_d$,去除冗余后的 $F_{cs}$更加突出了与图像内容密切相关的可区分特征。由于增强不同图像特征的区分度,因此能够较好提高零水印的唯一性。此外,$F_I$ 也保留了 $F_d$,从而保证了构造零水印的鲁棒性。

#### 1.1.3　抗攻击训练

为了提高零水印的鲁棒性,通过不断训练网络,使噪声图像提取的零水印近似接近于原始图像构造的零水印。由于零水印的构造来源于 DCRFENet 提取的图像特征,因此将原始图像 $I$ 与噪声图像 $N$ 的特征进行抗攻击训练。如图 1 所示,设 $I$ 和 $N$ 经过 DCRFENet 所提取的特征分别为 $F_I$ 和 $F_N$,采用均方误差(mean squared error, $MSE$)损失函数增加 $F_I$ 和 $F_N$ 之间的相似性:

$$L_{MSE} = MSE(F_I, F_N) = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(X_{i,j} - Y_{i,j})^2}{H \times W} , \tag{13}$$

式中,$X_{i,j}$为 $F_I$ 中位置为$(i, j)$的特征值,$Y_{i,j}$为 $F_N$ 中位置为$(i , j)$的特征值。

此外,为了使零水印能够抵抗不同类型的图像噪声攻击,对原始图像按照批量等概率的形式随机添加一种攻击的方式生成噪声图像,与原始图像生成的特征进行相似性训练。通过对 DCRFENet 的不断训练,不断降低 $L_{\text{MSE}}$损失,从而使 $F_I$ 和 $F_N$ 达到较高的相似度,提高了鲁棒特征的提取能力。当 $L_{\text{MSE}}$ 趋于稳定,则获得最终的 DCRFENet,该训练的 DCRFENet 用于提取构建零水印的鲁棒特征。

### 1.2　基于 DCRFENet 的零水印构造

在零水印构造方面,XIONG 等[7]采用图像块均值与所有图像块的均值进行大小比较,然而采用该方式构造的零水印唯一性较差。因此,本文采用对特征图进行分块,并将特征分块的每一个值与相应的均值进行比较构造零水印。设原始图像 $I$ 生成的零水印为$ZW$,原始水印为 $W_0$,它们的大小都为 $H \times W$,构造零水印的详细步骤如下:

步骤 1:将 $I$ 输入 DCRFENet,提取大小为 $H \times W$ 的特征图 $F_I$。

步骤 2:将 $F_I$ 分成大小为 $h \times w$ 互不重叠的分块 $P_k$,其中 $k$ 为每个分块的索引,对 $P_k$ 内的特征值进行二值化:

$$z_{i,j}^{k} = \begin{cases} 0, x_{i,j}^{k} < A_k \\ 1, x_{i,j}^{k} \geqslant A_k \end{cases} , \quad i = 1,2,\cdots,h; j = 1,2,\cdots w , \tag{14}$$

式中,$A_k$ 为 $P_k$ 的均值,重复对每一分块进行二值化,并组成二值图 $B_I$。

步骤 3:二值图 $B_I$ 与水印 $W_0$ 进行异或操作,生成零水印 $ZW$:

$$ZW = B_I \otimes W_0 。 \tag{15}$$

步骤 4:向时间戳权威机构申请时间戳与零水印 $ZW$ 绑定,并在知识产权数据库(intellectual property right database, IPRD)进行注册,从而实现对图像版权的保护。

### 1.3　基于 DCRFENet 的零水印提取

当接受图像时,图像有可能受到攻击,为验证该图像的版权,提取相应的零水印。设噪声图像为 $N$,提取零水印的过程与零水印构造过程相似,详细步骤如下:

步骤 1:将 $N$ 输入 DCRFENet,提取大小为 $H \times W$ 的特征图像 $F_N$。

步骤 2:将 $F_N$ 分成大小为 $h \times w$ 互不重叠的分块 $Q_k$,对 $Q_k$ 内的特征值进行二值化:

$$z_{i,j}^{k'} = \begin{cases} 0, x_{i,j}^{k'} < A'_k \\ 1, x_{i,j}^{k'} \geqslant A'_k \end{cases} , \quad i = 1,2,\cdots,h; j = 1,2,\cdots w , \tag{16}$$

式中,$A'_k$ 为 $Q_k$ 的均值。重复对每一分块进行二值化,并组成二值图 $B_N$。

步骤 3:将二值图 $B_N$ 与零水印 $ZW$ 进行异或操作,得到水印 $W_1$:

$$W_1 = B_N \otimes ZW 。 \tag{17}$$

步骤 4:若提取水印 $W_1$ 与原始水印 $W_0$ 相似,且时间戳通过认证,则认为版权申诉者具有该载体作品的合法版权。

## 2　实验结果及分析

本方法采用 Keras 网络框架,在 NVIDIA Ge-

Force RTX 2080 Ti 上执行。网络训练采用 CIFAR10 训练集[15]的 64 000 副彩色图像，随机梯度下降(stochastic gradient descent，SGD)作为优化器，以学习率为 0.000 1 进行迭代 15 次。为抵抗不同类型的攻击，在网络训练的每次迭代过程中，按照小批量(32 副图像为一组)等概率的形式随机选取一种强度较高的攻击(JPEG 压缩、均值滤波、椒盐噪声、旋转攻击、随机剪切)。

为了验证本方法的有效性，选取 512×512 的 8 副标准彩色图像作为测试图像，如图 3 所示。此外原始水印为随机生成的二值图。

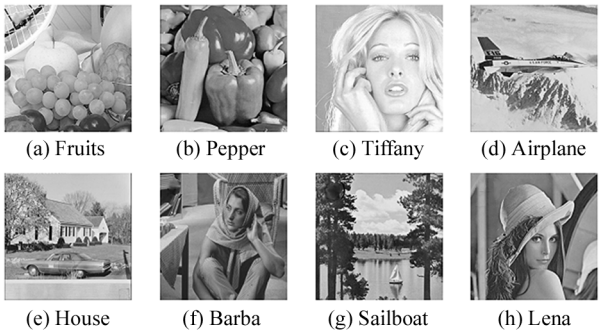


**图 3　标准测试图像**

**Fig. 3　Standard test images**

为了检测不同图像生成零水印的唯一性，采用归一化相关系数(normalized correlation coefficient，*NC* )客观评判零水印的相似性：

$$NC(ZW,ZW^*)=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}ZW(i,j)ZW^*(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}ZW^2(i,j)ZW^{*2}(i,j)}},\tag{18}$$

式中，$ZW$ 和 $ZW^*$ 分别表示不同图像构造的零水印。

此外，为了测试零水印的鲁棒性，采用比特误差率(bit error rate，*BER* )检测鲁棒性：

$$BER=\frac{N_e}{H\times W}\times 100\%,\tag{19}$$

式中，$N_e$ 表示提取错误水印比特的个数。

## 2.1　唯一性检测

为验证零水印的唯一性，对不同图像零水印进行相似度检测。如果不同图像零水印之间的 *NC* 值均小于 0.75，则认为其具有较好的唯一性[9]。表 1 为 8 副图像生成零水印相互之间的 *NC* 值，从表中可知 *NC* 的最大值为 0.666 6，表明不同图像构造的零水印相似度较低，即不同图像构造的零水印具有较好的唯一性。其主要原因为 DCRFENet 的特征消冗模块减少了冗余特征，提取了图像的有效特征，增强了图像特征的可区分性。此外，从CIFAR10测试集中随机选取200副图像进行唯一性检验，实验结果显示不同图像构造的零水印，其相互之间的 *NC* 值均小于 0.75，因此，再次证明本方法构造的零水印能够较好保护不同图像的版权。

**表 1　不同图像零水印之间的相似度(*NC*)**

**Tab. 1　Similarity between zero-watermarks generated by different images(*NC*)**

| | Fruits | Pepper | Tiffany | Airplane | Barba | Sailboat | Lena | House |
|---|---|---|---|---|---|---|---|---|
| Fruits | 1.000 0 | 0.476 0 | 0.473 9 | 0.448 2 | 0.511 9 | 0.495 4 | 0.479 3 | 0.424 5 |
| Pepper | 0.476 0 | 1.000 0 | 0.307 7 | 0.330 8 | 0.550 4 | 0.498 9 | 0.555 3 | 0.436 9 |
| Tiffany | 0.473 9 | 0.307 7 | 1.000 0 | 0.666 6 | 0.353 0 | 0.394 9 | 0.384 5 | 0.524 6 |
| Airplane | 0.448 2 | 0.330 8 | 0.666 6 | 1.000 0 | 0.342 3 | 0.426 3 | 0.392 8 | 0.528 5 |
| Barba | 0.511 9 | 0.550 4 | 0.353 0 | 0.342 3 | 1.000 0 | 0.477 7 | 0.488 4 | 0.431 9 |
| Sailboat | 0.495 4 | 0.498 9 | 0.394 9 | 0.426 3 | 0.477 7 | 1.000 0 | 0.494 1 | 0.518 9 |
| Lena | 0.479 3 | 0.555 3 | 0.384 5 | 0.392 8 | 0.488 4 | 0.494 1 | 1.000 0 | 0.415 5 |
| House | 0.424 5 | 0.436 9 | 0.524 6 | 0.528 5 | 0.431 9 | 0.518 9 | 0.415 5 | 1.000 0 |

## 2.2　鲁棒性检测

为测试基于 DCRFENet 的零水印方法对不同图像噪声攻击的抵抗能力，从 CIFAR10 测试集中随机选取 6 400 副图像，计算其 *BER* 的平均值作为本节鲁棒性衡量标准。首先，检验训练攻击的鲁棒性，包括不同强度的 JPEG 压缩、均值滤波、椒盐噪声、旋转攻击以及剪切攻击，如图 4 所示。图 4 显示了对于同一类型的攻击，当攻击强度增加时，相应的 *BER* 值会增加，但是都小于 0.09，该实验表明网络噪声攻击训练的有效性，对受训练的噪声攻击具有较强的鲁棒性。

其次，为了证明零水印方法的抗攻击泛化能力，检验对其他非训练攻击的鲁棒性，如图 5 所示。尽管高斯滤波、中值滤波、维纳滤波、缩放攻击以及高

斯噪声等图像攻击没有被训练,但是仍旧具有较强的鲁棒性,相应的 *BER* 都低于 0.05,其主要原因为 DCRFENet 提取了鲁棒特征,其次训练的攻击能够较好代表其他非训练攻击的内在特性。实验证明提出的零水印方法具有较好的抗攻击泛化性能。

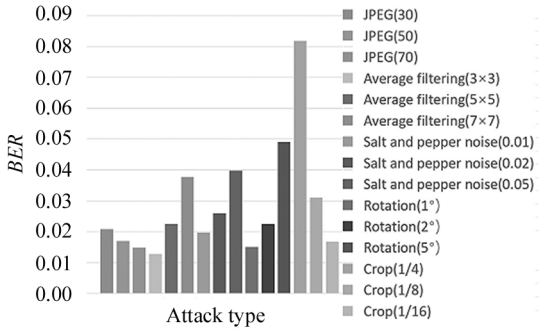


**图 4　抗训练攻击的鲁棒性**

**Fig. 4　Robustness on trained attacks**

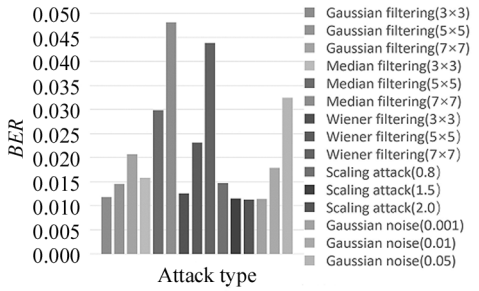


**图 5　抗非训练攻击的鲁棒性**

**Fig. 5　Robustness on untrained attacks**

此外,为了进一步证明本方法的普适性,从 CIFAR10 数据集[15]、COCO 数据集[16]、VOC 数据集[17]中分别随机选取 6 400 张测试图像进行鲁棒性测试。实验结果如表 2 所示,尽管采用 CIFAR10 进行训练,但对 VOC 与 COCO 数据集也均表现出较强的鲁棒性,其平均 *BER* 分别为 0.029 1,0.026 5 和 0.025 6。这主要是由于深度学习用于大量不同图像的特征学习,使得其提取的鲁棒特征能够适用于不同的图像类型,这也是优于传统零水印方法的鲁棒特征提取。

为了再次验证本零水印方法的较高鲁棒性能,与文献[4]—[9]进行比较。由于较难完全实现这些方法,为了尊重他们的实验结果,所比较的数据都是来源于相应发表的论文。图 6 为与文献[8]的鲁棒性能比较结果,计算测试图像为 Airplane、Fruits、Pepper、Sailboat、Lena 和 Barba 的均值。由图 6 可知,文献[8]在对抗低强度攻击时均表现出较好的鲁棒性,但其无法较好抵抗 JPEG 压缩(*QF* =10)、中值滤波(7×7)、椒盐噪声(0.05)等强度较高的攻击。然而,本方法不仅对低强度攻击表现出较强的鲁棒性,而且对于高强度攻击也表现出较好的鲁棒性。这主要由于训练 DCRFENet 网络时,采用 JPEG(*QF* =10)等高强度的攻击,从而使得对于较高强度的攻击也具有较好的抵抗力。此外,网络仅对 JPEG 压缩、均值滤波、椒盐噪声、旋转攻击等进行训练,然而在高斯滤波、中值滤波、高斯噪声、缩放攻击等非训练攻击,相应鲁棒性也好于文献[8],再次证明了本零水印方法具有较好的抗攻击普适性。

**表 2　不同数据集下的鲁棒性比较**

**Tab. 2　Robustness comparison under different datasets**

| Attack type | *BER* | | |
|---|---|---|---|
| | CIFAR10 | COCO | VOC |
| JPEG (10) | **0.028 7** | 0.033 7 | 0.031 8 |
| JPEG (50) | **0.017 0** | 0.018 3 | 0.017 7 |
| JPEG (90) | **0.012 8** | 0.016 4 | 0.016 0 |
| Gaussian filtering (5×5) | **0.014 5** | 0.015 6 | 0.015 2 |
| Gaussian filtering (9×9) | 0.0257 | 0.021 0 | **0.020 7** |
| Average filtering (5×5) | 0.0225 | 0.018 9 | **0.018 6** |
| Average filtering (9×9) | 0.056 2 | **0.037 0** | 0.037 2 |
| Median filtering (5×5) | 0.0299 | **0.027 5** | **0.027 5** |
| Median filtering (9×9) | 0.068 8 | **0.045 9** | 0.046 6 |
| Salt and pepper noise (0.01) | **0.019 7** | 0.025 2 | 0.023 5 |
| Salt and pepper noise (0.05) | **0.039 8** | 0.046 3 | 0.0431 |
| Rotation (1°) | **0.015 1** | 0.018 2 | 0.017 5 |
| Rotation (3°) | 0.030 8 | 0.031 9 | **0.030 3** |
| Scaling attack (0.5) | **0.015 4** | 0.016 6 | 0.016 2 |
| Scaling attack (1.2) | **0.012 8** | 0.016 0 | 0.015 4 |
| Gaussian noise (0.01) | 0.017 9 | 0.016 2 | **0.015 8** |
| Gaussian noise (0.05) | 0.032 4 | 0.036 2 | **0.015 8** |
| Wiener filtering (5×5) | 0.023 1 | 0.022 6 | **0.021 6** |
| Wiener filtering (9×9) | 0.069 4 | 0.059 1 | **0.055 1** |
| Average | 0.029 1 | 0.027 5 | **0.025 6** |

表 3 为与文献[5]、文献[6]的鲁棒性比较,*BER* 值为图 5 中 8 副图像的均值。由表 3 可知,尽管文献[5]和文献[6]对常见的图像处理表现出较好的鲁棒性,但其在旋转攻击上的鲁棒性表现较差,甚至无法抵抗低强度的旋转攻击。然而,本方法在抵抗低强度的旋转攻击方面具有较好的表现,随着旋转角度的增大,鲁棒性也会降低,但是相应 *BER* 仍旧低于 0.08,远远好于文献[5]和文献[6]。其主要原因为普通卷积具有一定的旋转不变性,但旋转不变性具有一定的角度限制。此外,相比文献[5],尽管抵抗高斯滤波(3×3)、椒盐噪声、缩放攻击的 *BER* 稍高,

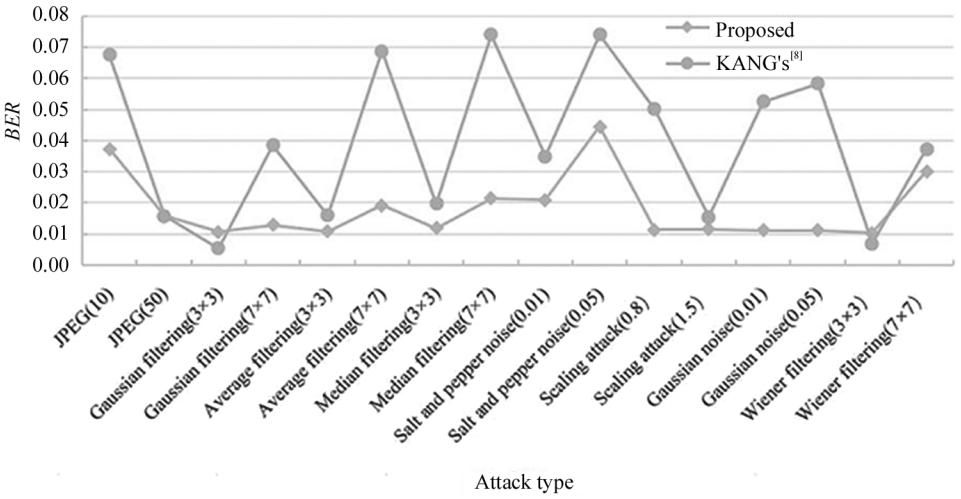


**图 6　不同攻击鲁棒性比较**

**Fig. 6　Robustness comparisons on different image attacks**

**表 3　8 副标准图像抗攻击比较**

**Tab. 3　Comparison of eight standard images against attacks**

| Attack type | Parameter | *BER* | | |
|---|---|---|---|---|
| | | Proposed | Ref. [5] | Ref. [6] |
| JPEG | 70 | **0.014 5** | 0.093 3 | 0.025 3 |
| | 50 | **0.016 7** | 0.145 4 | 0.039 2 |
| | 30 | **0.021 3** | 0.268 7 | 0.049 7 |
| Gaussian noise | 0.001 | **0.011 4** | 0.013 8 | 0.044 2 |
| | 0.005 | **0.011 3** | 0.057 6 | 0.072 0 |
| | 0.01 | **0.017 9** | 0.081 2 | 0.090 6 |
| Gaussian filtering | 3 × 3 | 0.010 6 | **0.008 5** | 0.012 5 |
| | 5 × 5 | **0.011 3** | 0.017 9 | 0.013 6 |
| Average filtering | 3 × 3 | **0.010 8** | 0.015 3 | 0.013 4 |
| | 5 × 5 | **0.013 8** | 0.054 2 | 0.017 3 |
| Salt and pepper noise | 0.01 | 0.022 7 | **0.013 6** | 0.059 1 |
| | 0.02 | 0.031 3 | **0.013 8** | 0.077 4 |
| Median filtering | 3 × 3 | **0.012 3** | 0.016 5 | 0.014 7 |
| | 5 × 5 | **0.016 6** | 0.046 7 | 0.018 3 |
| Rotation | 3° | **0.026 8** | 0.131 1 | 0.087 5 |
| | 5° | **0.042 9** | 0.191 2 | 0.119 8 |
| | 10° | **0.089 2** | 0.285 0 | 0.245 3 |
| Scaling attack | 0.5 | 0.011 7 | **0.009 1** | 0.014 9 |
| | 2 | 0.010 6 | **0.003 0** | 0.008 6 |
| Average | | **0.021 3** | 0.077 1 | 0.053 8 |

但本方法在抵抗 JPEG 压缩、高斯噪声、均值滤波、中值滤波等不同强度的攻击方面，相应的 *BER* 较低。与文献[5]相比，文献[6]的平均 *BER* 更低，具有更好的普适性。但文献[6]单一攻击的 *BER* 及所有攻击的平均 *BER* 都高于本方法，证明了基于DCRFENet的零水印方法既具有较强的鲁棒性又具有较好的普适性。

与单一攻击相比，提高零水印方法对混合攻击的抵抗能力具有更多的现实意义及更大的难度。为了检验对抗混合攻击的能力，采用 *NC* 作为鲁棒评价标准，与文献[4]、文献[7]以及文献[9]进行比较，实验结果如表 4 所示。文献[4]对混合攻击的鲁棒性表现较差，且 *NC* 均值低于 0.75，表明相应的零水印方法无法证明图像的版权。尽管文献[7]对抗维纳滤波(5×5)＋JPEG 压缩(10)和 JPEG 压缩(10)＋放大缩小 2 倍，以及文献[9]对抗中值滤波(5×5)＋JPEG压缩(10)和维纳滤波(5×5)＋椒盐噪声(0.3)的鲁棒性较好，但是综合考虑所有的组合攻击，相应的 *NC* 均值仍旧低于本方法。因此，混合攻击实验再次证明 DCRFENet 提取的图像特征具有较强鲁棒性，以及等概率随机攻击训练的有效性。综合上述，本方法在抵抗单一不同攻击(训练和非训练攻击)以及混合攻击方面具有较好的鲁棒性，相比其他方法也体现了较好的普适性。

为了进一步证明基于 DCRFENet 的零水印方法的有效性，与其他基于深度学习的零水印方法[13]进行比较。如表 5 所示，相比文献[13]，基于 DCFR-FENet 的零水印方法仅对椒盐噪声(0.01)的鲁棒性较弱，但抵抗其他不同攻击方面的鲁棒性较好。文献[13]采用过多的池化与卷积处理在一定程度上会丢失部分图像的主要信息，影响了鲁棒特征的提取。

表 4　抗混合攻击鲁棒性比较

Tab. 4　Comparison of robustness on hybrid attack

| Attack type | *NC* | | | |
|---|---|---|---|---|
| | Proposed | Ref. [4] | Ref. [7] | Ref. [9] |
| Median filtering (5×5) +Salt and pepper noise (0.3) | **0.910 3** | 0.578 4 | 0.890 1 | —— |
| Median filtering (5×5) + Gaussian noise (0.3) | **0.970 8** | 0.743 4 | 0.950 1 | 0.936 6 |
| Median filtering (5×5) +JPEG (10) | 0.960 0 | 0.867 6 | **0.976 2** | 0.981 5 |
| Wiener filtering (5×5) +Salt and pepper noise (0.3) | **0.899 3** | 0.583 6 | 0.892 6 | 0.934 7 |
| Wiener filtering (5×5) +Gaussian noise (0.3) | **0.976 8** | 0.764 1 | 0.956 6 | —— |
| Wiener filtering (5×5) +JPEG (10) | 0.965 6 | 0.883 5 | **0.981 9** | —— |
| JPEG (10) + Salt and pepper noise (0.3) | **0.906 2** | 0.586 6 | 0.892 9 | —— |
| JPEG (10) + Gaussian noise (0.3) | **0.979 1** | 0.761 0 | 0.955 1 | 0.936 1 |
| Rotation (2°) +JPEG (10) | **0.967 1** | 0.685 4 | 0.862 8 | 0.892 0 |
| JPEG (10) + Scaling attack (2) | 0.975 5 | 0.903 9 | **0.985 6** | 0.9694 |
| Average | **0.951 1** | 0.735 8 | 0.934 4 | 0.941 7 |

表 5　与文献[13]的鲁棒性比较

Tab. 5　Robustness comparison with Ref. [13]

| Attack type | *BER* | |
|---|---|---|
| | Proposed | Ref. [13] |
| JPEG (30) | **0.021 7** | 0.028 8 |
| JPEG (70) | **0.014 2** | 0.020 7 |
| Median filtering (3×3) | **0.013 7** | 0.021 4 |
| Median filtering (7×7) | **0.032 1** | 0.052 1 |
| Salt and pepper noise (0.01) | 0.025 2 | **0.021 7** |
| Salt and pepper noise (0.03) | **0.039 2** | 0.043 1 |
| Gaussian noise (0.002) | **0.009 3** | 0.032 4 |
| Gaussian noise (0.02) | **0.031 6** | 0.081 3 |
| Rotation (1°) | **0.018 2** | 0.042 4 |
| Rotation (5°) | **0.051 7** | 0.079 3 |
| JPEG (10) + Rotation (2°) | **0.041 2** | 0.082 1 |
| Median filtering (7×7) + Salt and pepper noise (0.02) | **0.068 1** | 0.080 4 |
| Average | **0.030 5** | 0.048 8 |

## 2.3　消融实验

为了进一步评估 DCRFENet 的密集连接模块和特征消冗模块对零水印的鲁棒性和唯一性的贡献，分别将特征图 $F_0$、$F_d$、$F_\alpha$ 经过 3×3 卷积输出，采用本方法相同的方式构造零水印，其相应的零水印方法分别命名为 BasicNet、DenseNet 和 RFENet。

表 6 为 4 种不同方法构造零水印的唯一性比较，实验数据为 CIFAR10 测试集的 200 副不同图像构造零水印的 *NC* 值在各个范围所占的比重。由表 6 可知，在 4 种零水印构造的方法中，BasicNet 与 DenseNet 构造的零水印 *NC* 范围整体偏高，甚至超过 0.75，而基于 DCRFENet 与 RFENet 构造的零水印均不超过 0.75，可见，不包含消冗模块零水印的唯一性较差。此外，由于 RFENet 仅比 DenseNet 多了特征消冗模块，但在唯一性上明显增强，因此从实验上证明了特征消冗模块能够明显增强零水印的唯一性。

表 6　零水印方法的唯一性比较

Tab. 6　Uniqueness comparison of zero water marking method

| Range of *NC* | *NC* ratio | | | |
|---|---|---|---|---|
| | Proposed | BasicNet | DenseNet | RFENet |
| Above 0.75 | 0.000 0 | 0.005 1 | 0.006 3 | 0.000 0 |
| Between 0.65 and 0.75 | 0.028 3 | 0.128 7 | 0.144 4 | 0.027 5 |
| Between 0.50 and 0.65 | 0.282 4 | 0.431 1 | 0.509 5 | 0.166 3 |
| Below 0.5 | 0.689 3 | 0.435 2 | 0.339 8 | 0.806 3 |

其次为了证明 DCRFENet 的密集连接模块对于鲁棒性的作用，4 种零水印方法也进行了抗攻击比较，如表 7 所示。相比 BasicNet、DenseNet 的平均 *BER* 明显更低，主要原因为 DenseNet 比 BasicNet 多了密集连接模块，从而证明密集连接模块对提高零水印的鲁棒性具有重要的作用。而 RFENet 的零水印鲁棒性低于本方法以及 DenseNet，主要原因也消除了一定的鲁棒特征，从而降低了相应水印的抗攻击能力。DenseNet 在抵抗不同图像攻击方面与本方法的 *BER* 值较接近，表明了本方法与 DenseNet 都具有较好的抗攻击能力，然而 DenseNet 的零水印唯一性方面较差。因此，综合考虑零水印的唯一性和鲁棒性，密集连接模块与特征消冗模块为 DCRFENet 重要组成部分，也表明基于 DCRFENet 的零水印能够为图像提供较好的版权保护功能。

表 7　4 种零水印网络模型的鲁棒性比较

Tab. 7　Robustness comparison of four zero watermarking network models

| Attack type | *BER* | | | |
|---|---|---|---|---|
| | Proposed | BasicNet | DenseNet | RFENet |
| JPEG (50) | **0.017 0** | 0.017 7 | **0.017 0** | 0.017 6 |
| JPEG (90) | **0.012 8** | 0.014 4 | 0.015 0 | 0.015 2 |
| Gaussian filtering (5×5) | **0.014 5** | 0.017 8 | 0.015 9 | 0.017 6 |
| Gaussian filtering (9×9) | **0.025 7** | 0.038 0 | 0.026 3 | 0.030 1 |
| Average filtering (5×5) | 0.022 5 | 0.032 8 | **0.022 3** | 0.026 6 |
| Average filtering (9×9) | 0.056 2 | 0.081 8 | **0.054 8** | 0.062 6 |
| Median filtering (5×5) | **0.029 9** | 0.037 3 | 0.030 2 | 0.032 0 |
| Median filtering (9×9) | **0.068 8** | 0.085 7 | 0.069 0 | 0.072 5 |
| Salt and pepper noise (0.01) | **0.019 7** | 0.023 0 | 0.020 6 | 0.021 0 |
| Salt and pepper noise (0.05) | 0.039 8 | 0.049 3 | **0.037 4** | 0.039 6 |
| Rotation (1°) | **0.015 1** | 0.019 7 | 0.017 5 | 0.018 7 |
| Rotation (3°) | **0.030 8** | 0.043 5 | 0.034 0 | 0.034 4 |
| Scaling attack (0.5) | **0.015 4** | 0.019 3 | 0.016 6 | 0.0183 |
| Scaling attack (1.2) | **0.012 8** | 0.013 6 | 0.015 5 | 0.015 1 |
| Gaussian noise (0.01) | **0.017 9** | 0.024 0 | 0.024 8 | 0.025 0 |
| Gaussian noise (0.05) | **0.032 4** | 0.035 2 | 0.034 8 | 0.036 5 |
| Wiener filtering (5×5) | **0.023 1** | 0.034 0 | 0.026 8 | 0.025 9 |
| Wiener filtering (9×9) | **0.069 4** | 0.093 9 | 0.075 2 | 0.072 0 |
| Average | **0.029 1** | 0.037 8 | 0.030 8 | 0.032 3 |

## 3　结　论

针对水印不可见性和鲁棒性的矛盾，设计了基于 DCRFENet 的零水印方法。首先，利用密集连接模块从各个网络层提取浅层和深层特征，提高零水印的鲁棒性。同时，采用特征间权重学习和特征内权重学习，提出特征消冗模块，增强零水印的唯一性。此外，对图像不同噪声进行抗攻击训练，从而提高零水印的鲁棒性。最后，基于训练的 DCRFENet，提取图像特征图，进行分块，然后利用分块均值与块内每一个特征值的大小关系构造零水印。实验结果表明，与现有零水印方法相比，提出的零水印方法具有较好的普适性，且对不同单一攻击以及混合攻击均具有较好的鲁棒性。

## 参考文献：


[1] LIU S W, DU Q Z, LONG H, et al. A robust audio watermarking algorithm based on DWT-DCT-SVD[J]. Journal of Optoelectronics · Laser, 2021, 32(9): 1015-1022.
刘思玮，杜庆治，龙华，等. 一种基于 DWT-DCT-SVD 的鲁棒性音频水印算法[J]. 光电子 · 激光，2021，32(9)：1015-1022.

[2] ZHENG Q M, CHEN Y R, LIN C. Contourlet watermarking algorithm based on geometric correction optimization[J]. Journal of Optoelectronics · Laser, 2022, 33(3): 330-336.
郑秋梅，陈亚茹，林超. 采用几何校正优化的轮廓波水印算法[J]. 光电子 · 激光，2022，33(3)：330-336.

[3] LI S Z, LI C, DENG X H. A tamper location and restoration approach for watermarking using image texture complexity[J]. Journal of Optoelectronics · Laser, 2019, 30(1): 44-51.
李淑芝，黎琛，邓小鸿. 基于纹理复杂度的图像篡改定位和恢复水印算法[J]. 光电子 · 激光，2019，30(1)：44-51.

[4] QU C B, YANG X T, YUAN D N. Zero-watermarking visual cryptography algorithm in the wavelet domain[J]. Chinese Journal of Image and Graphics, 2014, 19(3): 365-372.
曲长波，杨晓陶，袁铎宁. 小波域视觉密码零水印算法[J]. 中国图象图形学报，2014，19(3)：365-372.

[5] VELLAISARMY S, RAMESH V. Inversion attack resilient zero-watermarking scheme for medical image authentication[J]. IET Image Processing, 2014, 8(12): 718-727.

[6] ZOU B, DU J G, LIU X Y, et al. Distinguishable zero-watermarking scheme with similarity-based retrieval for digital rights management of fundus image[J]. Multimedia Tools and Applications, 2018, 77(21): 28685-28708.

[7] XIONG X G. A zero watermarking scheme with strong robustness in spatial domain[J]. Acta Automatica Sinica, 2018,44(1):160-175.
熊祥光.空域强鲁棒零水印方案[J].自动化学报,2018,44(1):160-175.

[8] KANG X, ZHAO F, CHEN Y, et al. Combining polar harmonic transforms and 2D compound chaotic map for distinguishable and robust color image zero-watermarking algorithm[J]. Journal of Visual Communication and Image Representation, 2020, 70:102804.

[9] KANG X, LIN G, CHEN Y, et al. Robust and secure zero-watermarking algorithm for color images based on majority voting pattern and hyper-chaotic encryption[J]. Multimedia Tools and Applications, 2020, 79(1):1169-1202.

[10] SHAO D G, HUANG J H, XU H. Segmentation of prostate image based on U-Net of multi-scale dilated separable convolution[J]. Journal of Optoelectronics · Laser, 2022, 33(5):554-560.
邵党国,黄俊辉,徐慧.基于多尺度空洞分离卷积的 U-Net 分割前列腺图像[J].光电子 · 激光,2022,33(5):554-560.

[11] EHSAEE S, JAMZAD M. Robust zero watermarking for still and similar images using a learning based contour detection[C]//Artificial Intelligence and Signal Processing, December 25-26, 2013, Tehran, Iran. Cham: Springer, 2014:13-22.

[12] HAN B, DU J, JIA Y, et al. Zero-watermarking algorithm for medical image based on VGG19 deep convolution neural network[J]. Journal of Healthcare Engineering, 2021, 2021:5551520.

[13] ATOANY F R, MARIKO N M, MANUEL C H, et al. A robust image zero-watermarking using convolutional neural networks[C]//2019 7th International Workshop on Biometrics and Forensics (IWBF), May 2-3, 2019, Cancun, Mexico. New York: IEEE, 2019:1-5.

[14] HUANG G, LIU Z, VAN D M L, et al. Densely connected convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, HI, USA. New York: IEEE, 2017:4700-4708.

[15] KRIZHEVSKY A, NAIR V, HINTON G. The CIFAR-10 [EB/OL]. (2014-04-09) [2022-06-29]. https://www.cs.toronto.edu/~kriz/cifar.html.

[16] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]//European Conference on Computer Vision-ECCV 2014, September 6-12, 2014, Zurich, Switzerland. Cham: Springer, 2014, 8693:740-755.

[17] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The PASCAL visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2):303-338.

**作者简介:**

**骆　挺**　(1980—),男,博士,教授,博士生导师,主要从事多媒体信息隐藏、多媒体视觉方面的研究.