文章编号:1672-3961(2018)03-0025-09 DOI:10.6040/j.issn.1672-3961.0.2017.408

基于 SVD 和 DCNN 的彩色图像多功能零水印算法

赵彦霞[1,2], 王熙照[1,3*]

(1. 河北大学管理学院, 河北 保定 071002; 2. 河北经贸大学信息技术学院, 河北 石家庄 050061;
3. 深圳大学计算机与软件学院, 广东 深圳 518060)

摘要:为了对彩色图像进行版权保护和篡改定位,提出一种基于奇异值分解(singular value decomposition, SVD)和深度卷积神经网络(deep convolutional neural network, DCNN)的彩色图像多功能零水印算法。将原始 RGB 彩色图像转换成 YCbCr 彩色图像,对原始图像的 Y、Cb、Cr 通道离散小波变换得到的系数矩阵进行奇异值分解,得到 DCNN 的输入矩阵,从 DCNN 输出层的输入矩阵中获取原始图像信息矩阵,生成零鲁棒水印图像。从 Y 通道小波变换得到的低频子带系数矩阵中获取原始图像信息矩阵,生成零半脆弱水印图像。试验结果证明,提出的算法不但有效,而且对强度较大的常见攻击有较好的抵抗能力。

关键词: 版权保护; 篡改定位;奇异值分解;深度卷积神经网络;多功能零水印;离散小波变换

中图分类号:TP391 **文献标志码**:A

引用格式:赵彦霞, 王熙照. 基于 SVD 和 DCNN 的彩色图像多功能零水印算法[J]. 山东大学学报(工学版),2018,48(3):25-33.
ZHAO Yanxia, WANG Xizhao. Multipurpose zero watermarking algorithm for color image based on SVD and DCNN[J]. Journal of Shandong University(Engineering Science), 2018, 48(3):25-33.

Multipurpose zero watermarking algorithm for color image based on SVD and DCNN

ZHAO Yanxia[1,2], WANG Xizhao[1,3*]

(1. College of Management, Hebei University, Baoding 071002, Hebei, China;
2. College of Information & Technology, Hebei University of Economics and Business, Shijiazhuang 050061, Hebei, China;
3. College of Computer Science & Software Engineering, Shenzhen University, Shenzhen 518060, Guangdong, China)

Abstract: A multipurpose zero watermarking algorithm for color image based on SVD (singular value decomposition) and DCNN (deep convolutional neural network) were proposed for the copyright protection and tamper location of color image. The original RGB color image was transformed into YCbCr color image. The Y channel, Cb channel and Cr channel were transformed by DWT (discrete wavelet transform), some matrices were got through decomposing the coefficient matrices by SVD and got the inputs matrices of DCNN. The information matrix of original image was got from the inputs matrix of output layer of DCNN and was used to generate zero robust watermarking image. The information matrix was got from the coefficient matrix of low frequency subband through the DWT of Y channel and was used to generate the zero semi-fragile watermarking image. The experimental results showed that the algorithm was not only efficient but also had good resistance to the strong common attacks.

Key words: copyright protection; tamper location; singular value decomposition; deep convolutional neural network; multipurpose zero watermarking; discrete wavelet transform

收稿日期:2017-05-17; **网络首发时间**:2018-06-19 16:37:24
网络首发地址:http://kns.cnki.net/kcms/detail/37.1391.T.20180619.0635.008.html
基金项目:国家自然科学基金资助项目(71371063, 61672205); 河北省应用基础研究计划重点基础研究资助项目(16960314D);河北省科技计划资助项目(15454704D);河北省人力资源社会保障科研合作课题资助项目(JRSHZ-2016-07038); 深圳市科技计划资助项目(JCYJ20150324140036825)

作者简介:赵彦霞(1970—),女,讲师,博士研究生,主要研究方向为不确定知识管理和机器学习,数字水印,信息处理等. E-mail:zyxa6@126.com
*通讯作者:王熙照(1963—),男,教授,博士,主要研究方向为机器学习,模式识别. E-mail:xizhaowang@ieee.org

0 引言

TIRKEL A Z 等人[1]提出了数字水印的概念。孙圣和等人[2]也指出,数字水印技术是通过在数字产品中嵌入秘密信息来保护数字产品的版权、证明产品的真实可靠性、跟踪盗版行为或提供产品的附加信息。研究人员已经提出了大量的数字水印算法[3-7]。向数字产品中加入水印,会改变数字产品,温泉等人[8]提出了零水印技术,不修改数字产品,而利用数字产品特征来构造水印(零水印)。

零水印技术是当前的研究热点之一,已经出现了大量针对视频[9-10]、音频[11-12]、文本[13-14]、图像[15-18]等数字产品的零水印算法。文献[8]利用高阶累积量提取的图像特征来构造零水印,但计算时间较长。文献[19]利用哈希值作为原始医疗图像的特征来构造零水印。文献[20]利用离散小波变换(DWT)和矩阵范数计算抗打印扫描攻击的不变特征值构造零水印。文献[21]将 BP 神经网络应用于零水印的构造中。文献[22]将宿主图像分成 4 * 4的子块,奇异值分解每一子块,利用 U 分量相邻系数大小关系保持不变的特征构造水印。但是文献[8]和文献[19-22]提出的算法,只构造了鲁棒水印,没有构造半脆弱水印来对待检测图像进行篡改检测。肖振久等人[23]利用超混沌技术,构造了医学图像篡改定位零水印,但还需要提高算法抵抗噪声攻击的能力。

零水印质量需要提取原始图像的关键特征来保障,且零水印算法应具有较强的抗攻击能力,考虑到这一点,本研究设计的算法中采用了 DWT、SVD 和 DCNN。与空域算法相比,变换域算法中水印鲁棒性更强。DWT 克服了离散傅里叶变换(discrete fourier transform, DFT)固定分辨率的缺点,也克服了离散余弦变换(discrete cosine transform, DCT)纯粹将空域变换到频率域和重构图像时容易出现马赛克现象的缺点。SVD 可以表示图像内在的代数特征,稳定性非常好。DCNN 通过卷积能够获取图像的关键特征。

1 算法相关技术介绍

1.1 奇异值分解

设矩阵 $\boldsymbol{A}_{m\times n}\in \mathbf{C}_r^{m\times n}(r>0)$,其中 $m\geqslant n$,则 $\boldsymbol{A}$ 的奇异值分解式[2]为:

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}=\sum_{i=1}^{n}\sigma_i u_i v_i^{\mathrm{T}}, \tag{1}$$

式中:$\boldsymbol{U}$、$\boldsymbol{V}$为正交矩阵;$\boldsymbol{S}$为非对角线元素都为0的矩阵;$\sigma_i=\sqrt{\lambda_i}(i=1,2,\cdots,n)$是 $\boldsymbol{S}$ 的对角线元素,是矩阵 $\boldsymbol{A}$ 的奇异值;$\lambda_i(i=1,2,\cdots,n)$是 $\boldsymbol{A}^H\boldsymbol{A}$ 的特征值[2]。

1.2 深度卷积神经网络

卷积神经网络(CNN)由 LECUN Y 等人[24]提出,是一种多层前馈网络。当网络有多个卷积层、下采样层时,就构成了深度卷积神经网络(DCNN),DCNN 一般由输入层、多个卷积层(C 层)和下采样层(S 层)、输出层构成。

DCNN 的训练分为 2 个阶段:前向传播阶段和反向传播阶段。

(1) 前向传播阶段。指由输入层向输出层的传播过程。在输入层输入样本,经过若干个卷积神经网络层,最后在输出层采用全连接输出特征图,将输出的特征图与期望的标签特征图比较,得到误差值,如果误差值在设定的允许误差范围内,则训练结束,否则进入反向传播阶段。

单层卷积神经网络的训练由卷积、非线性变换和下采样 3 个阶段完成。卷积和非线性阶段:输入卷积层的每个特征图与卷积核进行卷积,即采用局部权值共享方式连接,并加上偏置项,再经过非线性阶段(此阶段使用激活函数)。下采样阶段:对非线性阶段输入的特征图进行平均池化、最大池化或窗移等操作。

(2) 反向传播阶段。将误差值采用梯度下降的方式逐层向前传递,权值更新式(2)为:

$$\begin{aligned}\omega'^{l}_{ji}&=\omega^{l}_{ji}-\alpha\frac{\partial E}{\partial\omega^{l}_{ji}}=\omega^{l}_{ji}+\Delta\omega^{l}_{ji},\\ b'^{l}_{j}&=b^{l}_{j}-\alpha\frac{\partial E}{\partial b^{l}_{j}}=b^{l}_{j}+\Delta b^{l}_{j},\end{aligned} \tag{2}$$

式中:ω^{l}_{ji}和ω'^{l}_{ji}分别是更新前后的权重;E是误差值;α表示学习率;b^{l}_{j}和b'^{l}_{j}分别是更新前后的偏置。

2 基于 SVD 和 DCNN 的彩色图像多功能零水印算法

算法分为构造零水印图像算法和提取零水印图像算法。

2.1 构造零水印算法

构造零水印图像过程如图 1 所示。

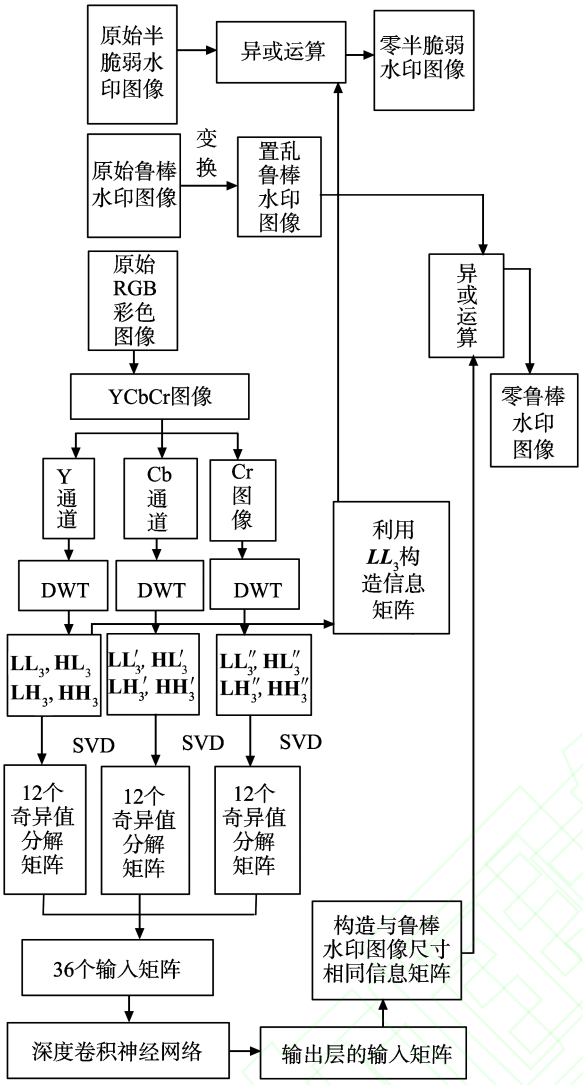


图 1　基于 SVD 和 DCNN 的彩色图像多功能构造零水印图像算法框图

Fig.1　The block diagram of multipurpose zero watermarking construction algorithm for color image based on SVD and DCNN

(1) 对原始彩色图像进行转换。读出原始图像,将其转换为 RCbCr 图像,并分解出 Y、Cb、Cr 通道。

(2) 读出鲁棒水印图像,并利用 Arnold 变换进行置乱。

(3) 读出半脆弱水印图像。

(4) 对原始图像的 Y、Cb、Cr 通道分别进行三尺度的“db1” 离散小波变换,并进行奇异值分解。Y、Cb、Cr 通道小波变换后第三层的子带小波系数矩阵分别为 $\mathbf{LL}_3,\mathbf{HL}_3,\mathbf{LH}_3,\mathbf{HH}_3$;$\mathbf{LL}'_3,\mathbf{HL}'_3,\mathbf{LH}'_3,\mathbf{HH}'_3$ 和 $\mathbf{LL}''_3,\mathbf{HL}''_3,\mathbf{LH}''_3,\mathbf{HH}''_3$。对这 12 个矩阵进行奇异值分解,共得到 36 个二维矩阵。

(5) 构造 DCNN 的输入矩阵和输出标签矩阵。将步骤(4) 中得到的 36 个矩阵转换成尺度为偶数的方阵。将 Y、Cb、Cr 通道中得到的方阵尺度最大值赋值给 m,创建 $m \times m \times 36$ 的特征矩阵 $\boldsymbol{X}$,其每个面为 $\boldsymbol{X}_n(n=1,2,\cdots,36)$,分别将步骤(4) 中得到的 36 个矩阵,赋值给 $\boldsymbol{X}_n$,就得到了输入特征矩阵。将 $\boldsymbol{X}_1$ 的前36列,每列之和除以 $\boldsymbol{X}_1$ 元素总个数,可得尺寸为 1 × 36 的输出标签矩阵 $\boldsymbol{T}$。

(6) 建立的 DCNN 如图 2 所示。

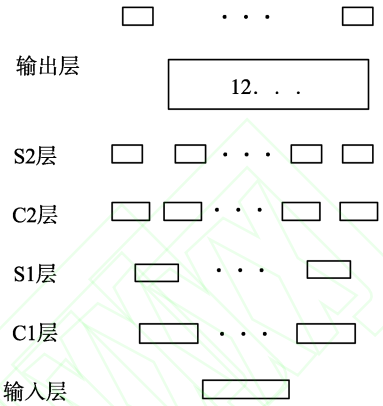


图 2　深度卷积神经网络结构示意图

Fig.2　Structural diagram of DCNN

图 2 的 DCNN 由输入层、卷积层 1(C1 层)、下采样层 1(S1 层)、卷积层 2(C2 层)、下采样层 2(S2 层)、输出层组成。卷积层所用卷积核的尺寸都为 5 × 5,C1 层和 C2 层分别输出 18 个和 36 个矩阵;S1 层和 S2 层分别将 C1 层和 C2 层传递的矩阵尺寸降为原来的 1/4。

(7) 对 DCNN 进行训练。训练分正向阶段和反向阶段。

① 正向阶段

(a) 输入层。输入 $\boldsymbol{X}$ 为一个大小为 $m \times m$ 的 36 个特征图,即为 $m \times m \times 36$ 的矩阵。

(b) 卷积层(C1 层和 C2 层)。对卷积层(l 层)的第 j 个输出特征图,计算公式[25] 为

$$Y_j^l = \sigma\Big(\sum_{i \in M_j^l} X_i^{l-1} * K_{ji}^l + b_j^l\Big), \tag{3}$$

式中:M_j^l 表示输入特征图的索引集合;K_{ji}^l 为大小为 5 × 5 的卷积核;“ * ” 代表卷积操作;σ 为 sigmoid 函数。C1 层和 C2 层分别输出 18 个 $(m-4)\times(m-4)\times 36$ 和 36 个 $(\frac{(m-4)}{2}-4)\times(\frac{(m-4)}{2}-4)\times 36$ 的特征图。

(c) 下采样层(S1 层和 S2 层)。计算下采样层第 j 个输出特征图[25]

$$Y_j^l = \beta_j^l \mathrm{down}(Y_j^{l-1}) + b_j^l, \tag{4}$$

式中:β_j^l 为常数权重 [0.25,0.25,0.25,0.25],down(·) 表示下采样。通过窗移,得到下采样层的输出特征图。S1 层和 S2 层分别得到 18 个

$\frac{(m-4)}{2}\times\frac{(m-4)}{2}\times 36$ 和 36 个 $\frac{(\frac{(m-4)}{2}-4)}{2}\times\frac{(\frac{(m-4)}{2}-4)}{2}\times 36$ 的特征图。

(d) 输出层。将 S2 层的输出转换成一维矩阵,作为输出层的输入矩阵,输出层的第 j 个单元输出 $o_j^l=\sigma(\sum_j^{n_o}\omega_{ji}^l\ Y_i^{l-1}+b_j^l)$。

② 反向传播训练

(a) 输出层的误差传递。设输出层为 L 层,输出层(全连接)的梯度

$$\frac{\partial E^n}{\partial \omega_{kj}}=-\delta_k^L\ y_j^{(L-1)},\tag{5}$$

式中:ω_{kj} 为 $(L-1)$ 层 j 单元与 L 层 k 单元的连接权重;$\delta_k^L=(o_k-t_k)o_k(1-o_k)$,为 L 层第 k 个单元的灵敏度;o_k 为 L 层第 k 个单元的输出;t_k 为 L 层第 k 个单元的期望输出。输出层的权重更新为

$$\Delta\omega_{kj}=\alpha\delta_k^L\ y_j^{(L-1)}=\alpha(o_k-t_k)o_k(1-o_k)y_j^{(L-1)},\tag{6}$$

式中:α 为学习率;$y_j^{(L-1)}$ 为 $(L-1)$ 层第 j 个单元的输出,也是 L 层第 j 个单元对应的输入。

(b) 采样层误差传递。S2 层的误差传递:S2 层的后一层是全连接层,其梯度计算式为

$$\frac{\partial E^n}{\partial \omega_{ji}}=\delta_j^{(L-1)}\ y_i^{(L-2)},\tag{7}$$

式中:$\delta_j^{(L-1)}$ 为 $(L-1)$ 层第 j 个单元的灵敏度。

(c) S1 层的误差传递。S1 层的后一层为卷积层(C2 层),其权重 β_j 和偏置 b_j 的梯度计算式[25] 为

$$\frac{\partial E^n}{\partial \beta_j}=\sum_{u,v}(\delta_j^{(L-3)}\ d_j^{(L-3)})_{uv},\ \frac{\partial E^n}{\partial b_j}=\sum_{u,v}(\delta_j^{(L-3)})_{uv},\tag{8}$$

式中:$\delta_j^{(L-3)}$ 为灵敏度,$d_j^{(L-3)}$ 为 C1 层输出特征图下采样后得到的特征图。$\delta_j^{(L-3)}$ 和 $d_j^{(L-3)}$ 的计算式[25] 分别为

$$\delta_j^{(L-3)}=\sigma'(x_j^{(L-3)})\circ \mathrm{conv2}(\delta_j^{(L-2)},\mathrm{rot180}(k_j^{(L-2)}),\text{‘full’}),\tag{9}$$

式中:“∘”表示每个元素相乘,$\delta_j^{(L-2)}$ 为灵敏度,$k_j^{(L-2)}$ 为权重。

$$d_j^{(L-3)}=\mathrm{down}(y_j^{(L-4)})。\tag{10}$$

本算法中,采样层的连接权重和偏置都采用常数,不需要调整,只是向其上层传递误差。

(d) 卷积层的误差传递。卷积层 l 的 δ_j^l 计算式[25] 为

$$\delta_j^l=\beta_j^{(l+1)}(\sigma'(\mathrm{net}_j^l)\cdot\mathrm{up}(\delta_j^{(l+1)})),\tag{11}$$

式中:$\beta_j^{(l+1)}$ 为下采样因子,是常数;$\sigma'(\mathrm{net}_j^l)$ 表示卷积层 l 的激活值 net_j^l 求导;$\mathrm{up}(\cdot)$ 表示上采样。

卷积层 l 的 k_{ji}^l 和 β_j^l 梯度计算式[25] 为

$$\frac{\partial E^n}{\partial k_{ji}^l}=\sum_{u,v}(\delta_j^l)_{uv}(p_i^{(l-1)})_{uv},\ \frac{\partial E^n}{\partial b_j^l}=\sum_{u,v}(\delta_j^l)_{uv}。\tag{12}$$

式中:$p_i^{(l-1)}$ 表示净激活 $\mathrm{net}_i^{(l-1)}$ 在卷积时与 k_{ji}^l 元素相乘的图像块。

(8) 获取用于构造零鲁棒水印图像的信息矩阵。取出 S2 层输出的矩阵 $\boldsymbol{F}$,取 $\boldsymbol{F}$ 中每一个 100 行 j 列($j=1,2,\cdots,36$) 为一组,记作第 i 组,计算其平均值。创建大小为 36 × 36 的矩阵 $\boldsymbol{G}$,如果 $\boldsymbol{F}$ 第 50 行 j 列的元素大于第 i 组元素平均值,则 $\boldsymbol{G}$ 的[i,j] 元素为 1,否则为 0,得到的 $\boldsymbol{G}$ 为用于构造零鲁棒水印的信息矩阵。

(9) 构造零鲁棒水印图像。将步骤(8) 获得的 $\boldsymbol{G}$ 和原始鲁棒水印进行异或运算,则产生零鲁棒水印图像。

(10) 获取用于构造半脆弱水印图像的信息矩阵。当 $\mathbf{LL}_3(i,j)\geqslant\mathbf{LL}_3((i+1),j)\geqslant\mathbf{LL}_3((i+2),j)$ 时,信息矩阵的元素为 1,否则为 0,形成与半脆弱水印尺寸相等的信息矩阵。

(11) 构造零半脆弱水印图像。将信息矩阵和原始半脆弱水印进行异或运算,产生零半脆弱水印图像。

2.2 提取零水印算法

提取零水印图像的过程如图 3 所示。

提取步骤如下:

(1) 读出零鲁棒水印图像。

(2) 读出零半脆弱水印图像。

(3) 读出待检测图像,转换成 YCbCr 图像,并分解出 Y^*、Cb^*、Cr^* 通道。

(4) 对待检测图像的 Y^*、Cb^*、Cr^* 通道分别进行三尺度的“db1” 离散小波变换,并进行奇异值分解。过程同 2.1 步骤(4)。

(5) 构造 DCNN 的输入矩阵和输出标签矩阵。过程同 2.1 步骤(5)。

(6) 建立如图 2 所示的 DCNN。

(7) 对 DCNN 进行训练。训练过程同 2.1 步骤(7)。

(8) 构造用于提取鲁棒水印图像的信息矩阵。过程同 2.1 步骤(8)。

(9) 提取鲁棒水印图像。将 2.2 步骤(8) 获取的信息矩阵和零鲁棒水印进行异或运算,提取出置乱的鲁棒水印图像,利用 Arnold 变换进行反置乱,则得到提取的鲁棒水印图像。

(10) 构造用于提取半脆弱水印图像的信息矩

阵。过程同 2.1 步骤(10)。

(11) 提取半脆弱水印图像。将 2.2 步骤(10)得到的信息矩阵和零半脆弱水印进行异或运算,则得到提取的半脆弱水印图像。

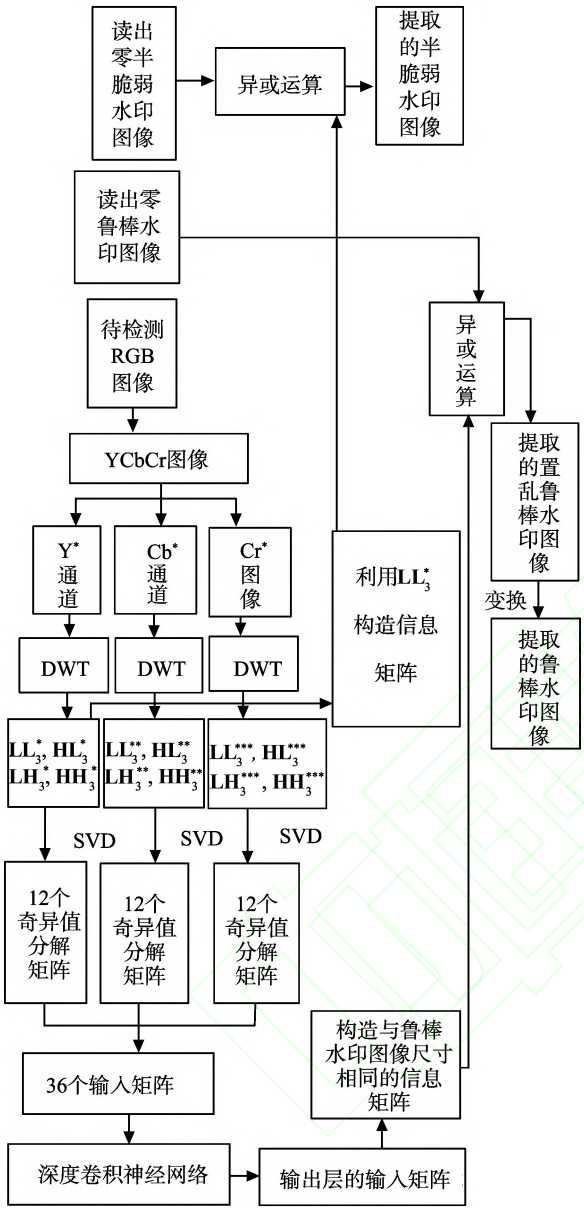


图 3　基于 SVD 和 DCNN 的彩色图像 多功能提取零水印图像算法框图

Fig.3　The block diagram of multipurpose zero watermarking extraction algorithm for color image based on SVD and DCNN

3　仿真试验与分析

3.1　仿真试验

算法采用 MATLAB 仿真实现。原始图像为 512 × 512 × 3 的 RGB 彩色图像,鲁棒水印和半脆弱水印都采用自制的像素灰度值只有 0 和 255 的图像,大小分别为 36 × 36 和 64 × 64,鲁棒水印置乱次数为 30 次。所采用的原始图像和水印图像如图 4 所示。

(a) 原始图像　(b) 鲁棒水印图像　(c) 半脆弱水印图像

图 4　原始图像和水印图像

Fig.4　Original image and watermarking image

3.1.1　构造和提取零水印图像试验

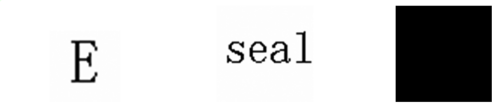
构造的零鲁棒水印图像和零半脆弱水印图像如图 5 所示。

(a) 零鲁棒水印图像　(b) 零半脆弱水印图像

图 5　零鲁棒水印图像和零半脆弱水印图像

Fig.5　Zero robust watermarking image and zero semi-fragile watermarking image

提取的水印图像与原始水印图像相似度用归一化互相关(normalised cross-correlation, NC)表示,不相似度用篡改评估函数(tamper assessment function, TAF)表示。原始图像未受攻击时,提取的水印图像和半脆弱水印图像篡改定位图如图 6 所示。提取的鲁棒水印图像和半脆弱水印图像的 NC 都为 1,提取的半脆弱水印图像的 TAF = 0。



(a) 提取的鲁棒水印图像　(b) 提取的半脆弱水印图像　(c) 篡改定位图

图 6　提取的水印图像和篡改定位图

Fig.6　Extracted watermarking image and extracted watermarking tamper location image

3.1.2　原始图像受攻击后试验

本研究对原始图像进行了剪切、拼贴、压缩、滤波、噪声、旋转等攻击试验。拼贴攻击分两种,拼贴 1 攻击是指从图像本身裁剪一部分进行拼贴攻击,拼贴 2 攻击是从其它图像裁剪一部分进行的拼贴攻击。图 7 为受到拼贴 1 攻击后的图像、提取出的水印图像及半脆弱水印图像篡改定位图,拼贴比例为待检测图像的 3.891 4%。

剪切、拼贴 1、拼贴 2、压缩、中值滤波、高斯噪声、椒盐噪声、旋转等常见攻击试验的结果如图 8 ~ 15 所示。图中“*”所在的曲线代表提取出的鲁棒水印的 NC,“○”所在曲线代表提取出的半脆弱水印的 NC,“+”所在曲线代表提取出的半脆弱水印的 TAF。

(a) 拼贴1攻击后图像

(b) 提取的鲁棒水印图像

(c) 提取的半脆弱水印图像

(d) 篡改定位图

图 7 拼贴 1 攻击后的图像、提取出的水印图像和篡改定位图

Fig.7 Image after the first collage attacks, watermarking image and tamper location image

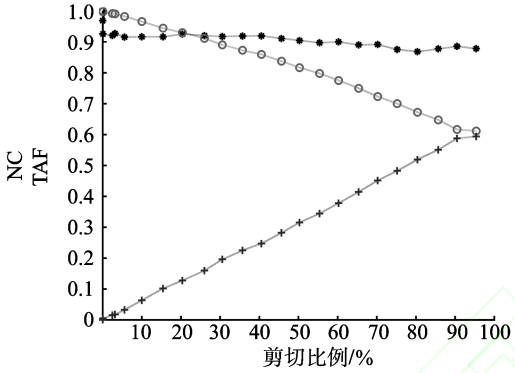


图 8 剪切攻击试验结果

Fig.8 Experimental results of cropping attacks

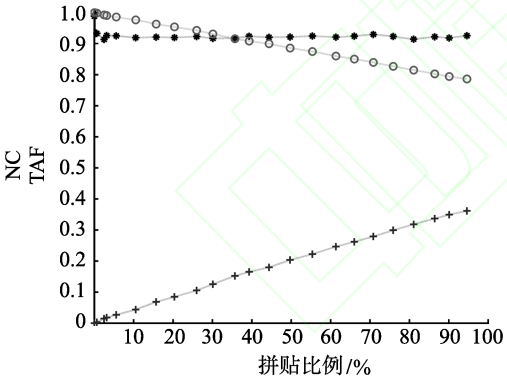


图 9 拼贴 1 攻击试验结果

Fig.9 Experimental results of the first collage attacks

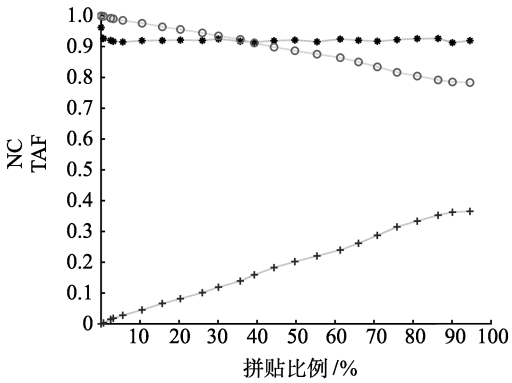


图 10 拼贴 2 攻击试验结果

Fig.10 Experimental results of the second collage attacks

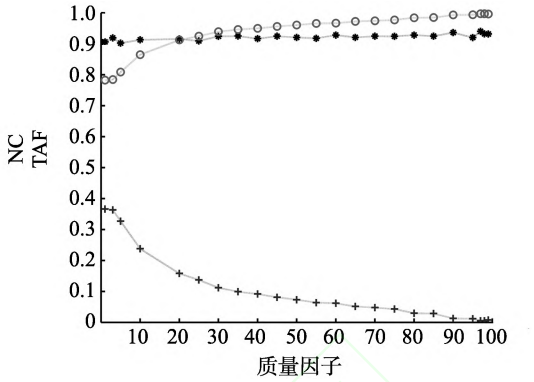


图 11 JPEG 压缩攻击试验结果

Fig.11 Experimental results of the JPEG compressing attacks

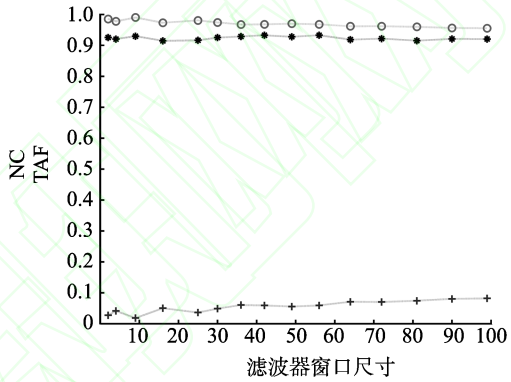


图 12 中值滤波攻击试验结果

Fig.12 Experimental results of the median filtering attacks

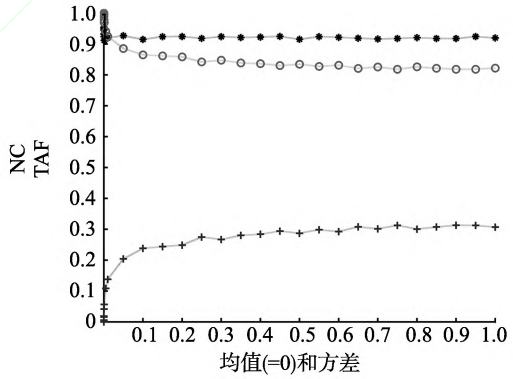


图 13 高斯噪声攻击试验结果

Fig.13 Experimental results of the gaussian noise attacks

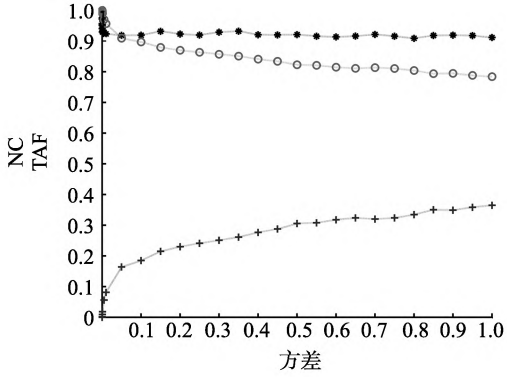


图 14 椒盐噪声攻击试验结果

Fig.14 Experimental results of the salt & pepper noise attacks

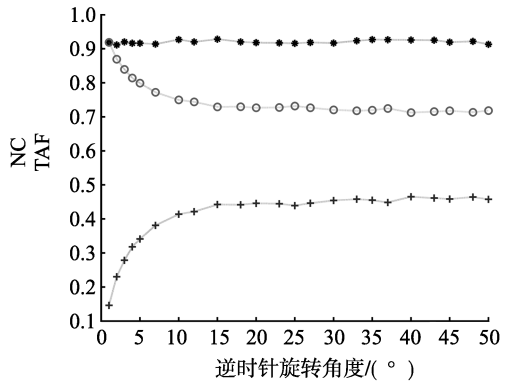


图 15　旋转攻击试验结果

Fig.15　Experimental results of the rotating attacks

3.2　与部分算法的比较

文献[21]提出了利用三层 BP 神经网络的数字图像零水印算法。所提出的算法使用的原始图像是 512×512 的 Lena 灰度图像,算法中只嵌入了一种自制的 32×32 的二值水印图像。将二维水印图像进行混沌置乱,再转换成一维数据。将宿主图像分成互不重叠的 3 * 3 子块,根据密钥,从宿主图像随机提取 N 个子块,记录每个子块中心点像素值和平均像素,建立 N 个子块到其平均像素的 BP 神经网络,根据每个子块中心点像素值和 BP 神经网络输出之间的关系获取二值序列,将此二值序列和混沌置乱后的一维水印数据进行异或运算得到构造出的二值零水印。受到攻击后,本研究算法和文献[21]算法提取的水印图像的 NC 如表 1 所示。

表 1　文献[21]和本研究算法水印 NC 比较

Table 1　Comparison of NC values of extracted watermarking images between the 21st reference and the diagram of this research

攻击类型	攻击强度	NC		
		水印图像(文献[21]基于分块奇异值分解的图像零水印算法)	鲁棒水印图像(本研究算法)	半脆弱水印图像(本研究算法)
中值滤波	3×3	0.902 0	0.929 8	0.990 2
高斯噪声	均值为 0,强度为 0.000 1	0.891 3	0.932 0	0.990 1
椒盐噪声	0.01	0.963 0	0.923 1	0.956 1
JPEG 压缩	质量因子为 90	0.979 7	0.936 0	0.993 1
	质量因子为 70	0.930 7	0.924 4	0.974 2
	质量因子为 50	0.903 2	0.920 4	0.960 4
缩放攻击	放大 2 倍	0.971 3	0.945 2	0.996 5

由表 1 可知,在受到中值滤波、高斯噪声、质量因子为 50 的 JPEG 压缩后,本研究算法提取出的鲁棒水印 NC 都大于文献[21]算法提取出的水印 NC。在受到表 1 所示的各种攻击后,除椒盐噪声攻击,本研究算法提取出的半脆弱水印 NC 都大于文献[21]算法提取出的水印 NC。从整体看,在受到表 1 所示的各种攻击后,本研究算法提取的鲁棒水印和半脆弱水印结合,抵抗攻击的能力几乎都强于文献[21]算法的抗攻击能力。

文献[23]提出一种基于超混沌的医学图像篡改定位零水印算法。算法所用的原始图像是 256×256 的 CT 图像,水印图像为 256×256 的二值图像。将原始图像最低有效位置零,将原始图像分成不重叠的子块,计算每个子块的平均值。根据子块每个像素值和子块平均值间的关系构成特征矩阵,将特征矩阵进行 Arnold 置乱,与超混沌加密的二值水印进行异或运算,从而构造零水印。受到攻击后,本研究算法和文献[23]算法提取的水印图像的 NC 如表 2 所示。

表 2　文献[23]和本研究算法提取的水印图像 NC 比较

Table 2　Comparison of NC values of extracted watermarking image between the 23rd reference and the diagram of this research

攻击类型	攻击强度	NC		
		水印图像(文献[23] 基于超混沌的医学图像零水印算法)	鲁棒水印图像(本研究算法)	半脆弱水印图像(本研究算法)
高斯噪声	0.002	0.706 7	0.924 0	0.818 0
	0.004	0.683 8	0.928 5	0.941 9
椒盐噪声	0.02	0.885 7	0.924 5	0.939 4
	0.04	0.821 5	0.922 2	0.917 4
剪切	1/8	0.989 5	0.920 4	0.956 2
	1/4	0.963 2	0.916 4	0.913 0

由表2可知,在受到表中的高斯噪声、椒盐噪声攻击时,本研究算法提取的鲁棒水印和半脆弱水印NC都大于文献[23]算法提取的水印NC。在受到表2中的剪切攻击后,本研究算法提取的鲁棒水印和半脆弱水印NC略低于文献[23]算法提取的水印NC。

综合表1、2中的试验结果,与文献[21,23]算法相比,本研究算法抵抗各种攻击能力大多优于文献[21,23]算法。

3.3 试验结果及分析

试验结果可以看出,本算法在原始图像未受攻击时,提取的鲁棒和半脆弱水印图像的NC都能达到1;当受到攻击时,鲁棒水印对剪切等攻击都有很好的抵抗能力。特别是由于采用了深度卷积网络,能够构造体现原始图像关键特征的零鲁棒水印图像,当攻击强度大时,鲁棒水印图像的相似度NC还能达到90%以上,能够很好地对原始图像进行版权保护。半脆弱水印对这些攻击能够较好地定位,也能对原始图像进行版权保护。不足之处是,在受到强度不太大的攻击时,提取鲁棒水印的NC有时不能达到98%以上。但是,鲁棒水印的这一不足,可以由半脆弱水印来补充,因为在攻击强度小时,提取的半脆弱图像一般能保持98%以上的NC,所以两种水印互补,使得本算法的综合性能较好。

4 结论

本研究提出一种基于SVD和DCNN的彩色图像多功能零水印算法。试验结果表明,该算法在受到常见攻击后,有较好的抗攻击能力,半脆弱水印在原始图像受到剪切等攻击时定位良好,特别是由于采用了DCNN,鲁棒水印对强度大的攻击的抵抗能力优于一般算法。算法可以用于医学图像、毕业证、身份证等证书图像的保护中,有较好的应用前景。

参考文献:

[1] TIRKEL A Z, RANKIN G A, SCHYNDEL R G, et al. Electronic watermark [C]//Digital Image Computing, Technology and Applications (DICTA), 1993. Sidney, Australia: Australian Pattern Recognition Society, 1993: 666-673.

[2] 孙圣和,陆哲明,牛夏牧,等. 数字水印技术及应用[M]. 北京: 科学出版社, 2004.

[3] AMIT M, NEELESH M. Adaptive lossless medical image watermarking algorithm based on DCT & DWT[J]. Procedia Computer Science, 2016, 78:88-94.

[4] 张君捧,张庆范,杨红娟.基于块特征和混沌序列的图像篡改检测与恢复[J]. 山东大学学报(工学版),2014,44(6):63-69.
ZHANG Junpeng, ZHANG Qingfan, YANG Hongjuan. Images tamper detection and recovery based on block features and chaotic sequence[J]. Journal of Shandong University(Engineering Science), 2014, 44(6):63-69.

[5] 王向阳, 杨红颖, 牛盼盼,等. 基于四元数指数矩的鲁棒彩色图像水印算法[J]. 计算机研究与发展,2016,53(3):651-665.
WANG Xiangyang, YANG Hongying, NIU Panpan, et al. Quaternion exponent moments based robust color image watermarking[J]. Journal of Computer Research and Development, 2016, 53(3):651-665.

[6] NGUYEN T S, CHANG C C, YANG X Q. A reversible image authentication scheme based on fragile watermarking in discrete wavelet transform domain[J]. International Journal of Electronics and Communications (AEU), 2016, 70(8):1055-1061.

[7] AMINI M, AHMAD M O, SWAMY M N S. Digital watermark extraction in wavelet domain using hidden Markov model[J]. Multimedia Tools and Applications, 2017, 76(3):3731-3749.

[8] 温泉, 孙锬锋, 王树勋. 零水印的概念与应用[J]. 电子学报, 2003, 31(2): 214-216.
WEN Quan, SUN Tanfeng, WANG Shuxun. Concept and application of zero-watermark[J]. Acta Electronica Sinica, 2003, 31(2): 214-216.

[9] ZHOU Z Y. Digital video zero-watermarking algorithm based on contourlet transform[J]. Microcomputer Information, 2010, 26(35):82-84.

[10] CHEN X S, BU G L, LI H T, et al. A video zero-watermark algorithm based on the contourlet transform [C]//The 3rd International Conference on Multimedia Technology (ICMT), 2013. Paris, France: Atlantis Press, 2013:216-223.

[11] CAI Y M, GUO W Q, DING H Y. An audio blind watermarking scheme based on DWT-SVD[J]. Journal of Software, 2013, 8(7):1801-1808.

[12] LEI M, YANG Y, LIU X M, et al. Audio zero-watermark scheme based on discrete cosine transform-discrete wavelet transform-singular value decomposition[J]. China Communications, 2016, 13(7):117-121.

[13] ZHU P, XIANG G L, SONG W N, et al. A text zero-watermarking algorithm based on Chinese phonetic alphabets[J]. Wuhan University Journal of Natural Sciences, 2016, 21(4): 277-282.

[14] JALIL Z, MIRZA A M, SABIR M. Content based zero-watermarking algorithm for authentication of text documents[J]. International Journal of Computer Science and

Information Security, 2010,7(2):248-255.

[15] KIM H D. CRT-based color image zero-watermarking on the DCT domain[J]. International Journal of Contents, 2015, 11(3):39-46.

[16] SINGH A, RAGHUVANSHI N, DUTTA M K, et al. An SVD based zero watermarking scheme for authentication of medical images for tele-medicine applications [C]//The 39th International Conference on Telecommunications and Signal Processing (TSP), 2016. Vienna, Austria: IEEE, 2016:511-514.

[17] LE H D, XU X Y, WANG Q, et al. Zero-watermarking for face image protection in database[J]. Journal of Internet Technology, 2016, 17(1):129-135.

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

[19] HAN B R, LI J B, LI Y J. Zero-watermarking algorithm for medical volume data based on difference hashing[J]. International Journal of Computers Communications & Control, 2015,10(2): 188-199.

[20] LI D, LIU Z, CUI L H. A zero-watermark scheme for identification photos based on QR code and visual cryptography[J]. International Journal of Security and Its Applications, 2016, 10(1): 203-214.

[21] 倪顾伟. 基于神经网络的数字水印算法的研究与实现[D]. 南京: 南京理工大学, 2012.
NI Guwei. Research and implementation of digital watermarking algorithm based on neural network[D]. Nanjing: Nanjing University of Science and Technology, 2012.

[22] 王雯霞. 一种基于奇异值分解的图像零水印新算法[J]. 计算机时代, 2010 (4): 8-10.
WANG Wenxia. A new image zero-watermarking algorithm based on SVD[J]. Computer Era, 2010(4): 8-10.

[23] 肖振久, 李南, 王永滨,等. 基于超混沌的医学图像篡改定位零水印算法[J]. 计算机工程与应用, 2017,53(7):115-120.
XIAO Zhenjiu, LI Nan, WANG Yongbin, et al. Zero watermarking scheme for medical image temper location based on hyper-chaos encryption[J]. Computer Engineering and Applications, 2017, 53(7):115-120.

[24] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[25] 谢剑斌,兴军亮,张立宁,等.视觉机器学习 20 讲[M]. 北京:清华大学出版社,2015.

(编辑:陈燕)

(上接第 24 页)

[19] GUTMANN M U, HYVÄRINEN A. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics[J]. Journal of Machine Learning Research, 2012, 13(2): 307-361.

[20] MNIH A, TEH Y W. A fast and simple algorithm for training neural probabilistic language models [C]// Proceedings of International Coference on International Conference on Machine Learning. Omnipress, Scotland: PMLR, 2012:419-426.

[21] MNIH A, KAVUKCUOGLU K. Learning word embeddings efficiently with noise-contrastive estimation [C]// Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe, USA: NIPS, 2013: 2265-2273.

[22] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China: ACM, 2001: 285-295.

[23] MNIH A, SALAKHUTDINOV R R. Probabilistic matrix factorization[C]// Proceedings of Advances in Neural Information Processing Systems. Whistler, Canada: NIPS, 2008: 1257-1264.

(编辑:陈燕)