



基于 SVD 和 SAE 的医学图像多功能零水印算法

孙洁丽¹, 赵彦霞^{1,2}, 周增慧³, 邢胜⁴

(1. 河北经贸大学信息技术学院, 河北 石家庄 050061;

2. 河北大学管理学院, 河北 保定 071002;

3. 河北经贸大学教务处, 河北 石家庄 050061;

4. 沧州师范学院计算机科学与工程学院, 河北 沧州 061001)

摘要: 针对医学图像的版权保护和篡改定位问题, 提出了一种多功能零水印算法。获取构造零顽健水印和半脆弱水印的信息矩阵, 并分别与原始顽健和半脆弱水印进行异或运算, 构造零顽健和零半脆弱水印图像。实验结果证明, 提取的水印效果较好, 算法能很好地抵抗常见攻击, 特别是抵抗强度较大的攻击的效果较好。

关键词: 医学图像; 多功能零水印; 离散小波变换; 奇异值分解; 堆叠自编码器

中图分类号: TN911.7

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017254

Multipurpose zero watermarking algorithm for medical image based on SVD and SAE

SUN Jieli¹, ZHAO Yanxia^{1,2}, ZHOU Zenghui³, XING Sheng⁴

1. College of Information and Technology, Hebei University of Economics and Business, Shijiazhuang 050061, China

2. College of Management, Hebei University, Baoding 071002, China

3. Educational Administration Office, Hebei University of Economics and Business, Shijiazhuang 050061, China

4. College of Computer Science and Engineering, Cangzhou Normal University, Cangzhou 061001, China

Abstract: A multipurpose zero watermarking algorithm was proposed according to the copyright protection and tamper location of medical image. The two information matrices were executed the XOR operation with the original robust watermarking image and the original semi-fragile watermarking image separately to generate zero robust watermarking image and zero semi-fragile watermarking image. The experimental results show that effects of extracted zero watermarking image are terrific, and the algorithm can resist the common attacks, particularly its resistance is effective for strong attacks.

Key words: medical image, multipurpose zero watermarking, discrete wavelet transform, singular value decomposition, stacked auto-encoder

收稿日期: 2017-06-27; 修回日期: 2017-08-05

基金项目: 国家自然科学基金资助项目 (No.71371063, No.61672205); 河北省应用基础研究计划重点基础研究项目 (No.16960314D); 河北省人力资源社会保障科研合作课题 (No.JRSHZ-2016-07038)

Foundation Items: The National Natural Science Foundation of China (No.71371063, No.61672205), The Key Basic Research Project of Applied Basic Research Program of Hebei Province (No.16960314D), The Science and Technology Cooperation Project of Hebei Provincial Department of Human Resources and Social Security (No.JRSHZ-2016-07038)

1 引言

医院每天都产生大量的医学图像等数字产品,对这些数字产品的保护也是对医院、医生和患者三方权益的保护。数字水印^[1]能够对数字产品进行版权保护和篡改定位,较好地保护这些产品。常用的数字水印是向原始数字产品嵌入数字水印来达到保护产品的目的,但提出的算法都会对数字产品进行修改,零水印技术^[2]被提出后,不再需要向数字产品中嵌入水印,而是提取原始数字产品的特征与数字水印一起构造零水印。目前,研究者已经提出了大量文本^[3,4]、音频^[5]、视频^[6,7]、图像^[8,9]等数字产品的零水印算法。

医学图像是医生给病人诊病的主要依据之一,当出现医患纠纷时,也可以作为医生治疗是否造成医疗事故的原始证据之一,构造顽健水印可以对医学图像进行版权保护,从而避免出现医学图像真假无法判定的问题。当有人将病灶部分修改或将病灶添加进医学图像时,希望能够通过技术手段检测到,因此,有必要构造半脆弱水印对被修改的医学图像进行篡改定位。另外,现代信息技术的快速发展,也为医疗机构之间通过医学资源共享提高诊病效率提供了可能,这也需要对医学图像进行版权保护和篡改定位。

对医学图像进行保护的水印算法研究已经成为一个研究热点,研究者提出了多种零数字水印算法。Seenivasagam 等人^[10]提出的零水印算法,利用混合 CT-SVD (contourlet transform-singular value decomposition) 域提取特征,将 QR 作为顽健水印,并利用胡氏不变矩构造零水印保证云端医学图像的安全性。Vellaisamy 等人^[11]提出了基于混合 CT-SVD 域的零水印算法,利用三角数生成函数和胡氏图像不变量抵抗反转攻击。Han 等人^[12]将三维医学图像进行三维离散小波和离散余弦变换,利用感知散列技术获取其本身特征向量,并利用 Legendre 混沌神经网络生成的混沌序列置乱

此特征向量。Dong 等人^[13]提取 DWT-DFT (discrete wavelet transform-discrete Fourier transform) 低频系数作为加密医学图像的特征向量,构造了顽健水印图像。但是,参考文献[10-13]都只构造了顽健水印对医学图像进行版权保护,没有对医学图像进行篡改定位。肖振久等人^[14]提出了利用超混沌技术的医学图像篡改定位零水印算法,但算法抵抗噪声攻击的能力还需要加强。

要构造高质量水印图像,需要获取能够代表原始图像的重要特征,并且构造的水印图像应该能够抵抗对原始图像的各种攻击。由于一般情况下,基于变换域构造的顽健水印比基于空域构造的水印的顽健性更强。奇异值分解 (singular value decomposition, SVD) 的稳定性好。属于深度神经网络的堆叠自编码器 (stacked auto-encoder, SAE) 能够获取图像的关键特征,并且还能够降低图像的维度,因此,提出的算法将 DWT、SVD 和 SAE 结合起来,构造了顽健水印和半脆弱水印,对医学图像进行版权保护和篡改定位。

2 相关技术

2.1 奇异值分解

$B \in C_r^{m \times n}$ ($r > 0$) 的大小为 $m \times n$, $m \geq n$, 则 B 的奇异值分解^[15]为式 (1)^[16]:

$$B = USV^T \quad (1)$$

其中, U 、 V 为正交矩阵, S 矩阵除对角线元素外,其他元素都为 0。

2.2 堆叠自编码器

堆叠自编码器^[17]是一个前馈神经网络,是由自动编码器 (auto-encoder, AE)^[18]堆叠而成的深度网络^[19]。堆叠自编码网络的训练分两步:自下而上的无监督学习,采用无标签样本,利用逐层贪婪学习,逐层对 AE 进行预训练;自上而下的有监督学习,对预训练得到的权重、偏置进行微调。

自动编码器是一个试图使输出与输入相同的简单神经网络。堆叠多个自动编码器可以构成深度网络。自动编码器的工作原理如图 1 所示。

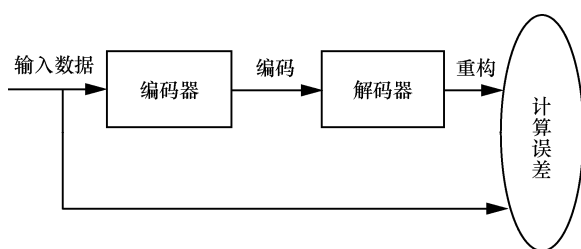


图1 自动编码器的工作原理

图1中，在网络训练期间，对每个训练样本进行训练后，会在输出层产生一个新的信号，网络学习的目的就是使输出信号与输入信号尽量相同。一个自动编码器的工作过程可分为3步：输入数据 x ，学习得到一个函数 f ，将输入数据编码成 $h=f(x)$ ；学习得到一个函数 g ，重构 h 为 $r=g(h)=g(f(x))$ ；利用损失函数 $L(r,x)$ 计算重构的 r 与输入数据 x 的误差，然后通过反向微调，最小化误差。一般情况下，中间层神经元数量远远少于输入层，在训练过程中得到的 h 的维度一般远远小于 x 的维度。如果将自动编码器与分类器结合，将 h 作为分类器的输入，将大大减少处理数据的信息量，从而大大提高计算效率。

3 基于SVD和SAE的医学图像多功能零水印算法

3.1 零水印算法构造流程

构造零水印图像的过程如图2所示，具体介绍如下。

步骤1 读出、转换和分解原始彩色图像。

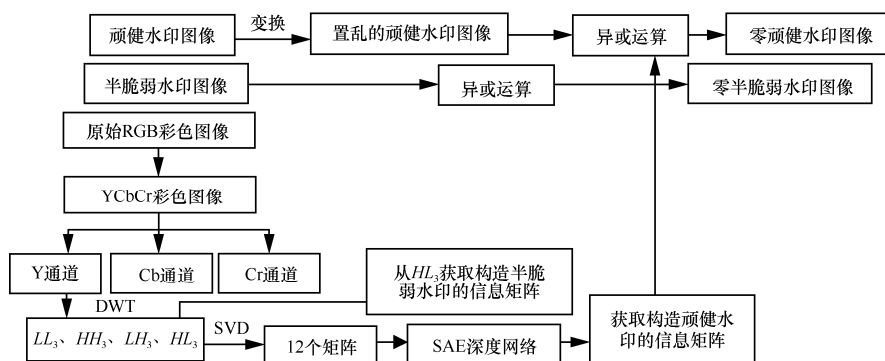


图2 零水印图像算法构造流程

读出原始图像，转换为YCbCr图像，分解出Y、Cb、Cr通道。

步骤2 读出和置乱顽健水印图像。读出顽健水印图像，利用Arnold变换对其进行置乱。设顽健水印图像尺寸大小为 $n_1 \times n_1$ 。

步骤3 读出半脆弱水印图像。设半脆弱水印图像尺寸大小为 $m_2 \times n_2$ 。

步骤4 对原始图像的Y通道进行小波分解和奇异值分解。利用三尺度的“db1”离散小波变换对原始图像的Y通道进行分解，然后再进行奇异值分解。Y通道小波变换后得到的第三层低频子带小波系数矩阵、对角方向、垂直方向和水平方向高频子带小波系数矩阵分别为 LL_3 、 HH_3 、 LH_3 、 HL_3 ，对这4个小波系数矩阵进行奇异值分解，得到12个矩阵。

步骤5 构造SAE深度网络的输入矩阵和输出标签矩阵。将步骤4中得到的12个矩阵的尺寸转换为偶数，将其最大值赋值给 m ，创建 $m \times m \times 12$ 的特征矩阵 l ，设其每个面为 $X_n(n=1,2,\dots,12)$ ，分别将步骤4中得到的12个矩阵，赋值给 $(l-1)$ ，就得到了输入特征矩阵。建立一个尺寸为 $8 \times 8 \times 12$ 的矩阵 Y_0 和一个尺寸为 $1 \times 8 \times 12$ 的矩阵 Y ，将 l 每个面的前8行×前8列矩阵赋值给矩阵 Y_0 每个面，将矩阵 Y_0 每个面的每列的和除以每个面元素总个数，所得的值赋值给 Y 的每个面。分别将 l 和 Y 转换成 $12 \times m \times m$ 和 $12 \times 1 \times 8$ 的二维矩阵，则得到的 l 和 Y 分别是SAE的输入矩阵和输出标签矩阵。

步骤 6 建立如图 3 所示的 SAE 深度网络。

图 3 所示为算法中用到的 SAE 深度网络结构, 由输入层 (V_0)、隐含层 1 (H_0 或 V_1)、隐含层 2 (H_1 或 V_2)、隐含层 3 (H_2) 和输出层构成, 各层之间神经元采用全连接, 本层神经元之间没有连接。输入层神经元个数为 $(m \times m)$, 设隐含层 1、隐含层 2、隐含层 3 的神经元个数分别为 N_1 、 N_2 、 N_3 , 隐含层 1、隐含层 2、隐含层 3 的神经元个数可根据情况确定, 但是需要满足 $(m \times m) > N_1 > N_2 > N_3$ 且 $(12 \times (N_3 + 1)) > (n_1 \times n_1)$, 输出层的神经元个数为 8。

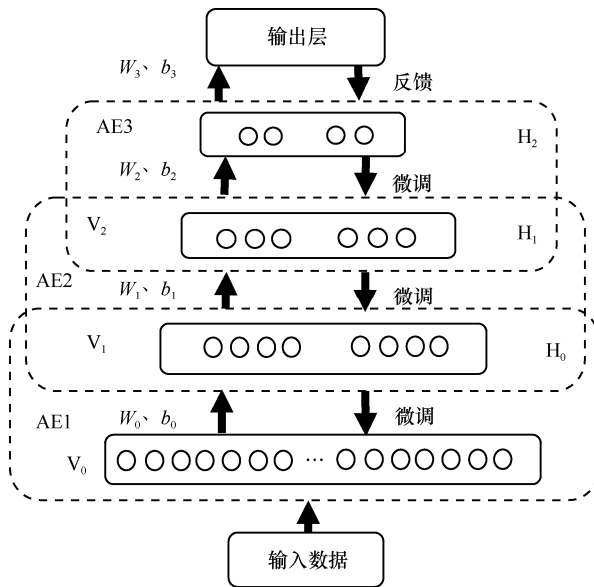


图 3 堆叠自编码器深度网络结构示意图

步骤 7 对 SAE 深度网络进行训练。将步骤 5 中产生的 SAE 的输入矩阵 x 作为堆叠自编码器隐含层 1 的输入值, 然后对 AE1、AE2 和 AE3 这 3 个自编码器进行预训练。

SAE 网络的训练分两步。

首先对 AE1、AE2 和 AE3 进行预训练。采用逐层贪婪学习方法训练每个自编码器, AE1 训练完成后, 其解码器的输出值 H_0 为 AE2 的输入值, AE2 解码器的输出值 H_1 为 AE3 的输入值。自编码器的工作示意如图 1 所示, 单个自编码器训练时, 需要经过编码和解码两个阶段。

堆叠自编码器 l 层第 k 个样本编码后的输出为 $y_k^{(l)}$ [17]:

$$y_k^{(l)} = \sigma(W_k^{(l)} x_k^{(l)} + b_k^{(l)}) \quad (2)$$

其中, $\sigma(\cdot)$ 为 Sigmoid 函数, $x_k^{(l)}$ 、 $y_k^{(l)}$ 、 $W_k^{(l)}$ 、 $b_k^{(l)}$ 分别为第 l 层第 k 个样本对应的编码器的输入、输出、连接权重和偏置。

堆叠自编码器 $(l+1)$ 层第 k 个样本解码重构后的输出为 $\hat{y}_k^{(l+1)}$ [17]:

$$\hat{y}_k^{(l+1)} = \sigma((W_k^{(l)})' y_k^{(l)} + b_k^{(l+1)}) \quad (3)$$

其中, $y_k^{(l)}$ 、 $\hat{y}_k^{(l+1)}$ 、 $(W_k^{(l)})'$ 、 $b_k^{(l+1)}$ 分别为第 $(l+1)$ 层第 k 个样本对应的解码器的输入、输出、连接权重和偏置。

再对预训练得到的权重、偏置进行微调。

将训练好的 3 个自编码器 AE1、AE2 和 AE3 按照堆叠的方式连接起来, 在 AE3 的 H2 之后加一层 softmax 分类器, 构成如图 3 所示的 SAE 深度网络, 根据输出误差, 利用反向传播算法对整个网络进行微调。反向传播算法的步骤为: 首先, 进行前向传导计算; 其次, 从输出层开始, 逐层向前传递误差, 对连接权重和偏置进行微调。

假设网络的层数为 N , 第 l 层的单元数量 (不含偏置) 用 s_l 表示。输入层第 i 个输入值 x_i 用 $d_i^{(l)}$ 表示, l 层的第 i 个单元的输用 $d_i^{(l+1)}$ 表示, $W_{ij}^{(l)}$ 表示 l 层第 j 个单元与 $(l+1)$ 层第 i 个单元的连接权重, $b_i^{(l)}$ 表示第 $(l+1)$ 层第 i 个单元的偏置值, $z_i^{(l)}$ 表示 l 层第 i 个单元输入加权和, $d_i^{(l)} = \sigma(z_i^{(l)})$, 其中, $\sigma(\cdot)$ 为 Sigmoid 函数。 $h_{w,b}(x)$ 为输出层的输出。

在前向传导阶段, SAE 第 $(l+1)$ 层的 $z_i^{(l+1)}$ 和 $d_i^{(l+1)}$ [20] 为:

$$z_i^{(l+1)} = \sum_{j=1}^{s_l} (W_{ij}^{(l)} d_j^{(l)} + b_i^{(l)}) \quad (4)$$

$$d_i^{(l+1)} = \sigma(z_i^{(l+1)}) \quad (5)$$

用向量表示各层数据后, SAE 第 $(l+1)$ 层的 $z^{(l+1)}$ 和 $d^{(l+1)}$ [20] 为:



$$z^{(l+1)} = W^{(l)} d^{(l)} + b^{(l)} \quad (6)$$

$$d^{(l+1)} = \sigma(z^{(l+1)}) \quad (7)$$

m 个训练样本的损失函数^[20]为:

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \| h_{W, b}(x^{(i)}) - y^{(i)} \|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{N-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \quad (8)$$

其中, $x^{(i)}$ 、 $y^{(i)}$ 分别为第 i 个训练样本的输入和输出。

在反向传播阶段, m 个样本损失函数的偏导为:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \right] + \lambda W_{ij}^{(l)} \quad (9)$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial b_i^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \quad (10)$$

用向量表示的输出层 (N) 的残差 $\delta^{(N)}$ 的计算式为:

$$\delta^{(N)} = -(y - d^{(N)}) \cdot \sigma'(z^{(N)}) \quad (11)$$

$l = N-1, N-2, \dots, 2$ 层残差的计算式为:

$$\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \cdot \sigma'(z^{(l)}) \quad (12)$$

偏导的计算式为:

$$\nabla_{W^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (d^{(l)})^T \quad (13)$$

$$\nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)} \quad (14)$$

l 层权值改变量 $\Delta W^{(l)}$ 和偏置改变量 $\Delta b^{(l)}$ 的计算式为:

$$\Delta W^{(l)} := \Delta W^{(l)} + \nabla_{W^{(l)}} J(W, b; x, y) \quad (15)$$

$$\Delta b^{(l)} := \Delta b^{(l)} + \nabla_{b^{(l)}} J(W, b; x, y) \quad (16)$$

权值更新计算式为:

$$W^{(l)} := W^{(l)} - \alpha \left[\frac{1}{m} \Delta W^{(l)} + \lambda W^{(l)} \right] \quad (17)$$

$$b^{(l)} := b^{(l)} - \alpha \left[\frac{1}{m} \Delta b^{(l)} \right] \quad (18)$$

其中, $\alpha = 1$, $\lambda = 0.5$ 。

步骤 8 获取用于构造顽健水印的原始图像

信息矩阵。取出输出层的输入矩阵 A , 创建一个尺寸为 $n_1 \times n_1$ 的元素值为 0 的矩阵 B , 从矩阵 A 隔 1 列取一个元素作为矩阵 B 的元素, 每取 n_1 个元素后, 矩阵 B 的行号加 1。创建尺寸为 $n_1 \times n_1$ 的元素值为 0 的矩阵 C 。当矩阵 B 的元素满足 $B(i, j) \leq B((i+1), j) \leq B((i+2), j)$ 时, 矩阵 C 的元素为 1, 否则为 0, 得到的矩阵 C 为用于构造零顽健水印的信息矩阵。

步骤 9 构造零顽健水印图像。将步骤 8 获得的矩阵 C 和置乱的原始顽健水印进行异或运算, 将产生的矩阵转换成像素值只有 0 和 255 的零顽健水印图像。

步骤 10 获取用于构造半脆弱水印的原始图像信息矩阵。利用 Y 通道离散小波变换后得到的水平方向系数矩阵 HL_3 得到信息矩阵, 当 $HL_3((i+2), j) \leq HL_3((i+1), j) \leq HL_3(i, j)$ 时, 信息矩阵的元素为 1, 否则为 0, 得到与半脆弱水印尺寸相等的信息矩阵。

步骤 11 构造零半脆弱水印图像。将步骤 10 得到的信息矩阵和原始半脆弱水印进行异或运算, 将产生的矩阵元素转换成像素灰度值只有 0 和 255 的零半脆弱水印图像。

3.2 零水印算法提取流程

提取零水印图像的过程如图 4 所示, 具体介绍如下。

步骤 1 读出零顽健水印图像。

步骤 2 读出零半脆弱水印图像。

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

步骤 4 对待检测图像的 Y* 通道进行离散小波变换, 并进行奇异值分解。对待检测图像的 Y* 通道进行三尺度的“db1”离散小波变换, Y* 通道小波变换后得到的第三层的低频子带小波系数矩阵、对角方向、垂直方向和水平方向高频子带小波系数矩阵分别为 LL_3^* 、 HH_3^* 、 LH_3^* 、 HL_3^* , 对这 4 个小波系数矩阵进行奇异值分解, 得到 12 个矩阵。

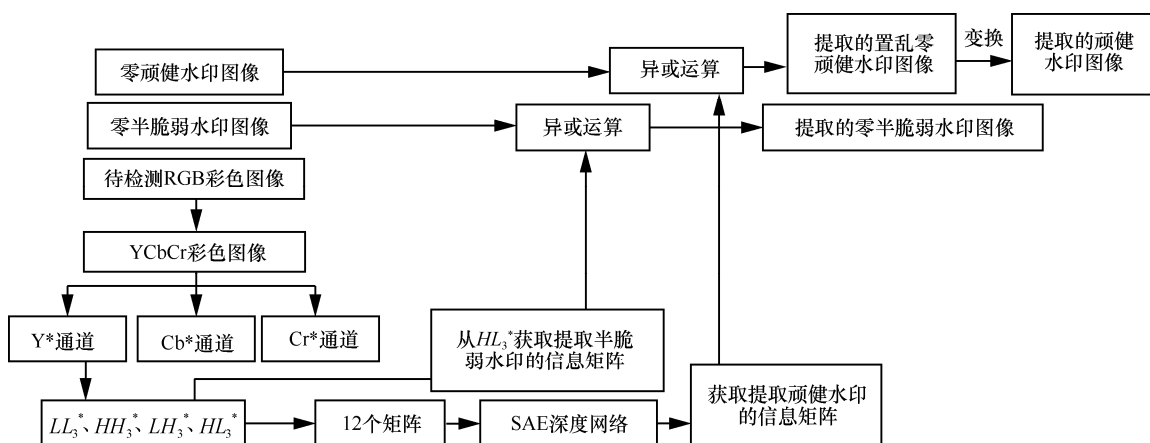


图4 零水印图像算法提取流程

步骤5 构造SAE深度网络的输入矩阵和输出标签矩阵。过程同第3.1节的步骤5。

步骤6 建立如图4所示的SAE深度网络。

步骤7 对SAE深度网络进行训练。训练过程同第3.1节的步骤7。

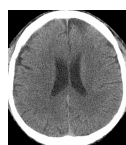
步骤8 构造用于提取顽健水印图像的信息矩阵。过程同第3.1节的步骤8。

步骤9 提取顽健水印图像。将零顽健水印和第3.2节的步骤8得到的信息矩阵进行异或运算，提取出像素值为0和1的置乱顽健水印图像，转换成像素值为0和255的图像，对其进行Arnold变换，则获得提取的顽健水印图像。

步骤10 获取用来提取半脆弱水印图像的信息矩阵。过程同第3.1节的步骤10。

步骤11 提取半脆弱水印图像。将零半脆弱水印和第3.2节的步骤10得到的信息矩阵进行异或运算，则获得提取的半脆弱水印图像。

Arnold变换的次数为60次。第3.1节的步骤6的 $m \times m = 52 \times 52 = 2704$, $N_1 = 1000$, $N_2 = 500$, $N_3 = 205$, 输出层神经元个数为8。所用的原始图像和水印图像如图5所示。



(a) 原始图像

M

(b) 顽健水印图像

WA

(c) 半脆弱水印图像

图5 原始图像和水印图像

4.1.1 零水印图像构造和提取实验

构造的零顽健和零半脆弱水印图像如图6所示。



(a) 零顽健水印图像



(b) 零半脆弱水印图像

图6 零顽健水印图像和零半脆弱水印图像

4 仿真实验和结果分析

4.1 仿真实验

算法采用MATLAB仿真实现。原始图像为 $408 \times 354 \times 3$ （单位：像素）的CT图像，顽健水印和半脆弱水印图像均采用自制图像，图像的像素灰度值只有0和255，尺寸分别为 36×36 （单位：像素）和 51×45 （单位：像素），顽健水印进行

分别用归一化互相关（normalised cross-correlation, NC）和篡改评估函数（tamper assessment function, TAF）表示提取的水印图像与原始水印图像的相似度和不相似度。在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像如图7所示。提取的顽健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。

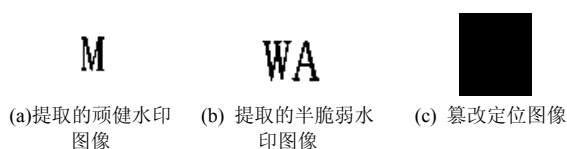


图7 提取的水印图像和篡改定位图像

4.1.2 原始图像受攻击实验

实验中,对待检测图像进行了多种攻击。对待检测图像进行剪切比例为 27.972 3%的剪切攻击后的图像、提取出的水印图像和半脆弱水印图像的篡改定位图像如图 8 所示。图 8 中,提取的顽健水印的 $NC=0.9890$, 半脆弱水印的 $NC=0.8726$, 半脆弱水印的 $TAF=0.2092$ 。

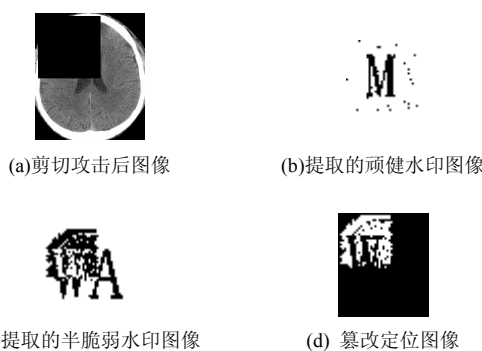


图8 剪切攻击后的图像、提取出的水印图像及篡改定位图像

如图 9~图 16 所示,分别为原始图像受到剪切、拼贴、中值滤波、旋转、JPEG 压缩、高斯噪声、椒盐噪声和缩放等常见攻击实验后,提取的顽健水印图像 NC、半脆弱水印图像 NC 和半脆弱水印图像 TAF 值的测量值。

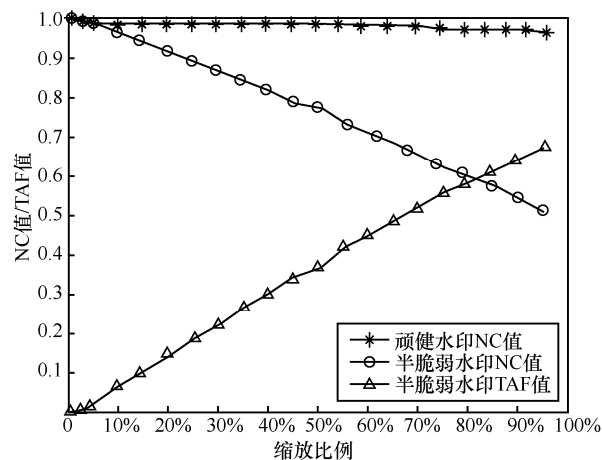


图9 剪切攻击实验结果

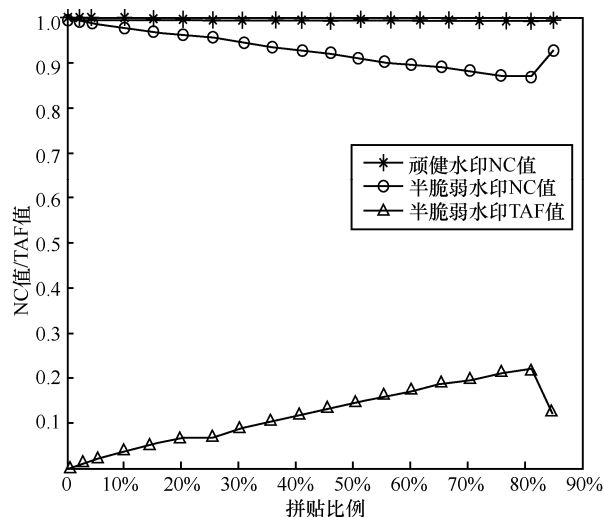


图10 拼贴攻击实验结果

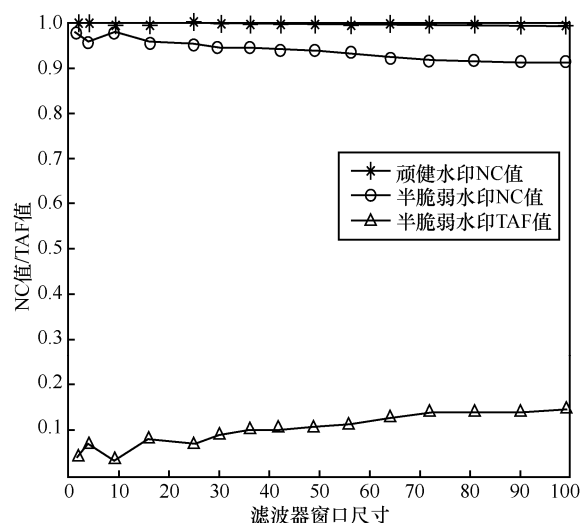


图11 中值滤波攻击实验结果

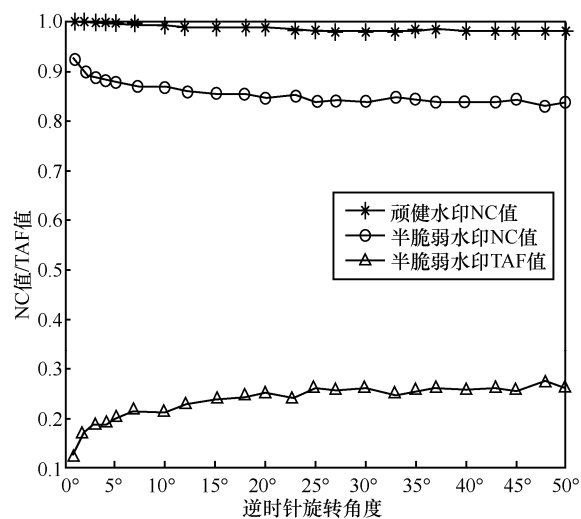


图12 旋转攻击实验结果

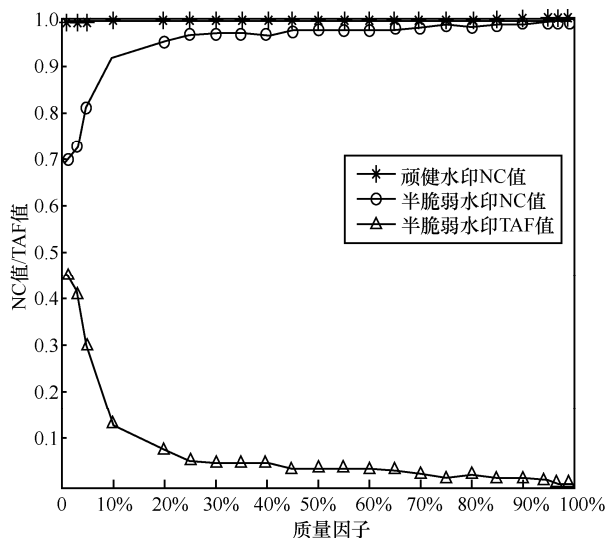


图 13 JPEG 压缩攻击实验结果

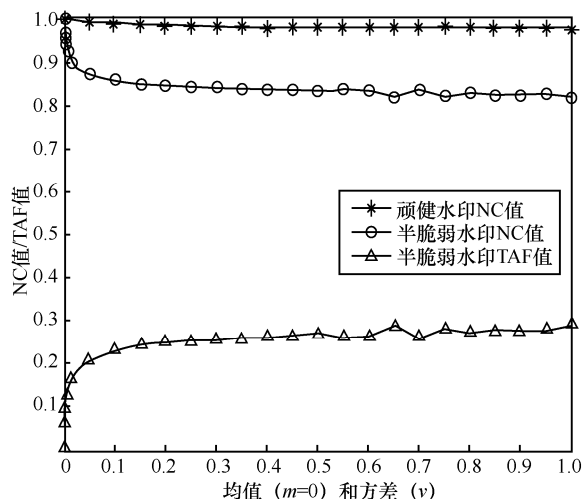


图 14 高斯噪声攻击实验结果

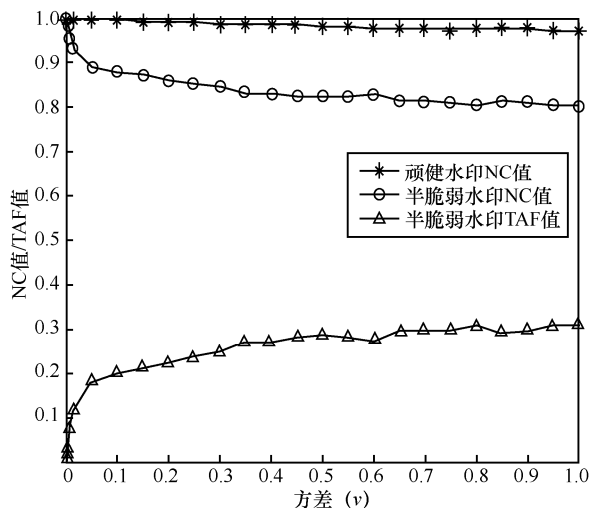


图 15 椒盐噪声攻击实验结果

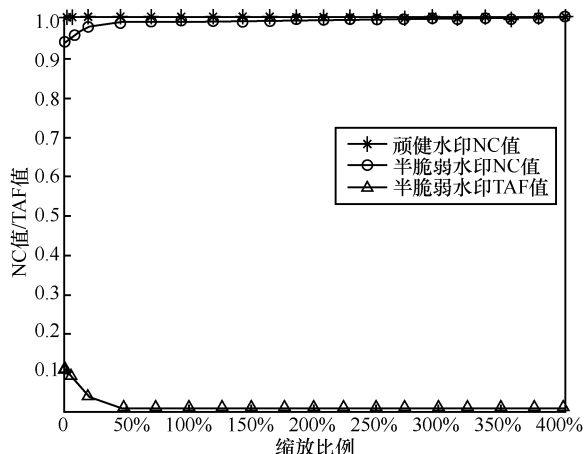


图 16 缩放攻击实验结果

4.2 实验结果及分析

由实验结果可看出, 本算法在原始图像未受攻击时, 提取的顽健水印和半脆弱水印图像的 NC 值都为 1。在原始图像受到多种攻击时, 提取的顽健水印图像 NC 值在 0.964 9 以上, 说明提取的水印有较好的顽健性。另外, 半脆弱水印也能够很好地对多种攻击进行篡改定位。

对于 RGB 彩色图像, 亮度信息代表图像的主要信息, 由于彩色图像的 Y 通道代表原始图像的亮度信息, Cb、Cr 通道代表原始图像的色度信息。为了减少计算的数据量, 提出的算法选择在 Y 通道中提取原始图像关键特征信息。在构造零顽健水印时, 考虑到 SVD 能代表图像的代数特征, 在算法中采用了 SVD, 又进一步利用 SAE 能够提取图像关键特征的优点, 获取原始图像的关键特征。另外, DWT 的应用也可以增加顽健水印的顽健性。在构造半脆弱水印时, 选择了 DWT 得到的水平方向高频子带小波系数矩阵 HL_3 构造半脆弱水印, 也能很好地对受到攻击后的图像进行篡改定位。

许多水印算法中仿真实验使用的图像, 对应的矩阵都是方阵, 这里实验所用的 CT 图像不是方阵, 是医院真实的图像。实验结果表明, 所提出的算法对于真实的医学图像具有很好的效果, 具有很好的实用价值。



5 结束语

提出了一种基于 SVD 和 SAE 的医学图像多功能零水印算法。实验结果表明, 顽健水印的顽健性很好, 半脆弱水印能够很好地对篡改进行定位, 算法能够抵抗多种常见攻击。提出的算法, 创新性地将 SAE 能够较好提取图像特征的优势应用到构造顽健水印的过程中, 使构造的顽健水印对多种攻击, 特别是强度大的攻击都有很好的顽健性, 研究成果可以应用于证书、电子签字等多种数字产品的保护中, 有很好的实用价值。进一步对算法进行研究, 结合领域的实际情况, 将算法推广应用于更多的领域, 可以保证这些领域数字产品的安全, 对这些领域的发展做出贡献。

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), Dec 8-10,1993, Sidney, Australia. Sydney: Australian Pattern Recognition Society, 1993: 666-673.
- [2] 温泉, 孙铤锋, 王树勋. 零水印的概念与应用[J]. 电子学报, 2003, 31(2): 214-216.
WEN Q, SUN T F, WANG S X. Concept and application of zero-watermark[J]. Acta Electronica Sinica, 2003, 31(2): 214-216.
- [3] 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(4): 277-282.
- [4] 周新民, 鲁芳. 基于汉字结构知识的文本认证零水印算法[J]. 吉林大学学报(理学版), 2016, 54(5): 1071-1076.
ZHOU X M, LU F. Zero-watermarking algorithm for text authentication based on structure knowledge of Chinese characters[J]. Journal of Jilin University (Science Edition), 2016, 54(5): 1071-1076.
- [5] 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.
- [6] 王静, 郁梅, 李文峰, 等. 抗量化转码的 HEVC 视频流零水印算法[J]. 光电工程, 2016, 43(10): 77-83.
- [7] WANG J, YU M, LI W F, et al. HEVC-based video stream zero-watermarking algorithm with robustness to re-quantization transcoding[J]. Opto-Electronic Engineering, 2016, 43(10): 77-83.
- [7] XU Q S, LU J F, PENG X X, et al. A video zero-watermarking algorithm based on text detection[C]//2015 16th IEEE International Conference on Communication Technology(ICCT2015), Oct 18-20, 2015, Hangzhou, China. New Jersey: IEEE Press, 2015: 328-333.
- [8] 徐莲, 李京兵. 基于 DWT-PCA 和感知哈希的鲁棒零水印算法[J]. 电视技术, 2016, 40(1): 25-30.
XU L, LI J B. Robust zero-watermarking algorithm based on DWT-PCA and perceptual hashing[J]. Video Engineering, 2016, 40(1): 25-30.
- [9] 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.
- [10] SEENIVASAGAM V, VELUMANI R. A QR code based zero-watermarking scheme for authentication of medical images in teleradiology cloud[J]. Computational & Mathematical Methods in Medicine, 2013(4): 1-16.
- [11] VELLASAMY S, RAMESH V. Inversion attack resilient zero-watermarking scheme for medical image authentication[J]. IET Image Processing, 2014, 8(12): 718-727.
- [12] HAN B R, CAI L S, LI W F. Zero-watermarking algorithm for medical volume data based on legendre chaotic neural network and perceptual hashing[J]. International Journal of Grid Distribution Computing, 2015, 8(1): 201-212.
- [13] DONG J T, LI J B. A robust zero-watermarking algorithm for encrypted medical images in the DWT-DFT encrypted domain[J]. Innovation in Medicine and Healthcare, 2016, 60(4): 197-208.
- [14] 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, 2015, 53(7): 115-120.
XIAO Z J, LI N, WANG Y B, et al. Zero watermarking scheme for medical image temper location based on hyper-chaos encryption[J]. Computer Engineering and Applications, 2015, 53(7): 115-120.
- [15] 薛栋, 周亚训, 金炜. 版权保护和内容认证的全盲双功能数字水印算法[J]. 电信科学, 2017, 33(2): 79-89.
XUE D, ZHOU Y X, JIN W. Bifunctional watermarking for copyright protection and content authentication with perfectly blind detection[J]. Telecommunications Science, 2017, 33(2): 79-89.
- [16] GHAZY R A, ABBAS A M, AI-ZUBI N, et al. Block-based

- SVD image watermarking in spatial and transform domains[J]. International Journal of Electronics, 2015, 102(7): 1091-1113.
- [17] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy layer-wise training of deep networks[C]//Advances in Neural Information Processing Systems 19(NIPS2006), Dec 4-7, 2006, Vancouver, Canada. Cambridge: MIT Press, 2007: 153-160.
- [18] BENGIO Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [19] YANN L C, YOSHUA B, GEOFFREY H. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [20] NG A. Sparse autoencoder[EB/OL]. [2017-06-18]. <https://www.mendeley.com/research-papers/sparse-autoencoder/>.

[作者简介]



孙洁丽（1969-），女，博士，河北经贸大学教授，主要研究方向为信息处理、数据挖掘。



赵彦霞（1970-），女，河北大学管理学院博士生，河北经贸大学讲师，主要研究方向为数字水印、不确定知识管理与机器学习、信息处理。



周增慧（1975-），男，河北经贸大学工程师，主要研究方向为信息管理、数据挖掘。



邢胜（1982-），男，博士，沧州师范学院讲师，主要研究方向为不确定知识管理与机器学习。