计算机科学
COMPUTER SCIENCE

一种抗屏摄攻击的DCT域深度水印方法

黄昌喜, 赵成鑫, 姜骁腾, 凌贺飞, 刘辉

引用本文

黄昌喜, 赵成鑫, 姜骁腾, 凌贺飞, 刘辉. 一种抗屏摄攻击的DCT域深度水印方法[J]. 计算机科学, 2024, 51(2): 343-351.

HUANG Changxi, ZHAO Chengxin, JIANG Xiaoteng, LING Hefei, LIU Hui. Screen-shooting Resilient DCT Domain Watermarking Method Based on Deep Learning [J]. Computer Science, 2024, 51(2): 343-351.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221200121

一种抗屏摄攻击的 DCT 域深度水印方法

黄昌喜　赵成鑫　姜骁腾　凌贺飞　刘　辉
华中科技大学计算机科学与技术学院　武汉 430074
(hcx2022@hust.edu.cn)

摘　要　数字水印技术在多媒体保护方面发挥着巨大的作用,实际应用需求的变更推动了数字水印技术的发展。目前,基于深度学习的水印技术在鲁棒性上有了较大的提升,但水印的嵌入基本在空域进行,载体图像的失真仍然比较明显。此外,现有方法在面对摄屏攻击时效果不佳。为解决上述问题,提出了一种抗屏摄攻击的 DCT 域深度水印方法。该模型由 DCT 层、编码器、解码器和屏摄模拟层组成。DCT 层将图像的 Y 分量转换为 DCT 域,然后编码器通过端到端训练修改 DCT 系数,将秘密消息嵌入到图像中。这种频域嵌入方法使得水印信息能够分布到图像的整个空间,从而减少了失真效应。此外,还提出了一个噪声层,用于模拟屏摄过程中特殊的摩尔纹和反光效果。训练过程分为两个阶段:在第一阶段,编码器和解码器进行端到端的训练;而在第二阶段,屏摄模拟层和传统的失真攻击被用来增强水印图像,然后使用失真水印图像来进一步优化解码器。大量的实验结果表明,该模型具有较高的透明度和鲁棒性,并且在屏摄鲁棒性方面优于其他方法。

关键词:数字水印;深度学习;DCT 变换;不可感知性;屏摄鲁棒性
中图分类号　TP391

Screen-shooting Resilient DCT Domain Watermarking Method Based on Deep Learning

HUANG Changxi,ZHAO Chengxin,JIANG Xiaoteng,LING Hefei and LIU Hui
School of Computer Science and Technology,Huazhong University of Science and Technology,Wuhan 430074,China

Abstract　Digital watermarking technology plays an important role in multimedia protection,and the various demands for practical applications promotes the development of digital watermarking technology. Recently,the robustness of the deep learning-based watermarking model has been greatly improved,but the embedding process is mostly carried out in the spatial domain,and this causes obvious distortions to original images. In addition,existing methods do not work well under the screen-shooting attack. To solve the above problems,this paper proposes a deep learning-based DCT domain watermarking method which is robust to the screen-shooting attack. This model consists of a DCT layer,an encoder,a decoder,and a screen shoot simulation layer. The DCT layer converts the Y component of images into the DCT domain,then the encoder embeds secret messages into the image by modifying the DCT coefficients through end-to-end training. This embedding method in the frequency domain makes the watermark information to be distributed to the whole space of images so that the distortion effect is reduced. Furthermore,we propose a noise layer to simulate moiré and light reflection effects,which are common distortions in the screen-shooting attack. The training process is splitted into two stages. In the first stage,the encoder and decoder are trained end-to-end. While in the second stage,the screen-shooting simulation layer and traditional distortion attacks are used to augment the watermarked image,then we use the distorted watermarked image to furtheroptimize the decoder. Extensive experimental results show that the proposed model has high transparency and robustness,and is superior to other methods in screen robustness.

Keywords　Digital watermark,Deep learning,DCT Transform,Imperceptibility,Screen-shooting robustness

1　引言

作为信息隐藏技术的分支,数字水印技术指在载体文件中以不可见的形式嵌入信息,并在载密图像经过含有失真的信道传输后,仍能从中恢复原始信息。作为版权保护和泄密溯源的有利工具,数字水印需要具有透明性和鲁棒[1-2]。透明

到稿日期:2022-12-20　返修日期:2023-05-14
基金项目:国家自然科学基金(61972169);国家重点研发计划(2019QY(Y)0202,2022YFB2601802);湖北省重点研发计划(2022BAA046,2022BAA042);武汉基础研究知识创新项目(2020010601012182);中国博士后科学基金(2022M711251)
This work was supported by the National Natural Science Foundation of China(61972169),National Key Research and Development Program of China(2019QY(Y)0202,2022YFB2601802),Major Scientific and Technological Project of Hubei Province(2022BAA046,2022BAA042),Research Programme on Applied Fundamentals and Frontier Technologies of Wuhan(2020010601012182) and China Postdoctoral Science Foundation (2022M711251).
通信作者:凌贺飞(lhefei@hust.edu.cn)

性指水印的嵌入不能改变原有图片的视觉感受。鲁棒性则要求水印系统能够抵抗各种类型的失真攻击。随着“屏读”时代的来临，人们越来越多地从显示器中获取信息，屏摄过程成为信息传输和泄密的重要途径，而抗屏摄是鲁棒水印的一个难点。不管是传统水印方案还是基于深度学习的水印技术，目前都无法较好地解决该问题，因为屏摄过程中存在特殊的镜头失真、摩尔纹失真和光照失真[3]，而此类失真在保留图像语义的同时，极大程度上破坏了嵌入的水印信号。因此，抗屏摄水印的研究充满挑战又具有重大意义。

传统的数字水印方法通常寻找载体图像中的不重要区域，并根据具体失真类型，将秘密信息嵌入到变换域中的失真不变特征上。但传统方案往往需要对特定的失真类型做具体设计，要求大量的人工先验知识，且存在泛化性差的问题。近些年，深度学习技术的引入，进一步推动了数字水印技术的发展。Zhu 等[4]提出的“编码器-噪声层-解码器”的深度水印框架被认为是提升水印模型鲁棒性最有效的方案之一，该框架通过端到端训练方式获得了更为出色的鲁棒性。大部分基于深度学习的水印模型采用的视觉一致性损失往往直接在像素级别上做约束，而基于空域的修改方案难以避免地表现为对轮廓等特征的修改，使得水印对原始图像的修改较为明显，载密图像的视觉失真严重。

针对该问题，本文将传统水印方案中变换域的经验做法与深度水印技术相结合，提出了一种基于 DCT 域的深度水印模型。传统 DCT 域水印方法通过修改 DCT 系数以嵌入水印，将水印信息分布到了载体图像的整个区域，人眼难以察觉到水印修改带来的变化，并且抵抗 JPEG 压缩等各种噪声攻击的能力较好。但传统 DCT 域的水印方法由于是人为设计，因此存在鲁棒性和透明性的权衡困难的问题。结合神经网络的特征学习能力和频域水印的优秀性能可以进一步弥补目前水印方法的不足。具体来讲，本文在编码器之前加入 DCT 变换层，将图像的 Y 分量进行 DCT 变换，使得编码器自学习频域嵌入模式。

为提高水印模型对该类失真的鲁棒性，还提出了一种摩尔纹失真和反光噪声模拟方案。提出的水印模型将模拟过程加入端到端训练，具有优于其他方案的抗屏摄的鲁棒性能。

本文的主要贡献如下：

1）提出了一种 DCT 域深度水印方法。与传统深度水印直接嵌入在载体图像空域不同，本文利用 DCT 域系数稳定特性，将图像 Y 分量的 DCT 系数作为编码空间，由编码器网络自适应修改不同频域系数实现水印嵌入。

2）设计屏摄攻击模拟噪声层。本文利用图像处理技术对摄屏过程中常见的摩尔纹和反光效应进行模拟，对载密图像加入屏摄攻击效果，对解码器进行了针对屏摄攻击的鲁棒性增强训练。

3）提出了一种抗屏摄攻击的 DCT 域深度水印方法。本文采用二阶段训练的方式，第一阶段训练无失真条件下获取高保真载密图像的编解码器，第二阶段对解码器进行屏摄鲁棒性增强训练。实验证明，提出的方法与同类方法相比具有更高的视觉质量和对抗电子信道失真和屏摄失真的鲁棒性。

2 相关工作

2.1 数字水印技术

传统的数字水印主要基于空域和变换域来对图像这一载体上的某些比特位进行修改。最经典的空域方法是最低有效位（Least Significant Bit，LSB）方法[5]，该方法通过修改图片中有效位的最低位来实现信息存储，但载密图像的简单失真会破坏嵌入的水印信息。随后，研究者发现图像的变换域存在针对特定失真的不变特性，因此变换域水印技术将图像先变换到频率域，通过改变图像的频域系数进行水印嵌入，然后通过反变换得到载密图像。目前研究较为成熟的是基于离散余弦变换[6]（Discrete Cosine Transform，DCT）和离散小波变换[7]（Discrete Wavelet Transform，DWT）的数字水印算法。传统算法更多的是依赖设计者的经验。这类方法由于其针对性的设计，往往只对某一类图像有着较好的效果，在其他情况下，效果就会大打折扣。近年来，深度学习发展火热且表现出了优越的性能，人们开始考虑将水印算法和深度学习相结合。

最初，神经网络用于传统水印系统中的一个模块，如 Fang 等[8]提出利用神经网络模型来增强嵌入的水印模板特征。而后，Zhong 等[9]提出了全部由神经网络构成的水印模型，其中网络通过约束由全连接层组成的不变特征提取层来保证水印模型的鲁棒性。Zhu 等[4]提出的 HiDDeN 模型展现出了远超传统水印方法的性能，其提出“编码器（Encoder）-噪声层（Noiser）-解码器（Decoder），END”的水印框架，该框架成为了深度水印模型的研究主流，如图 1 所示。

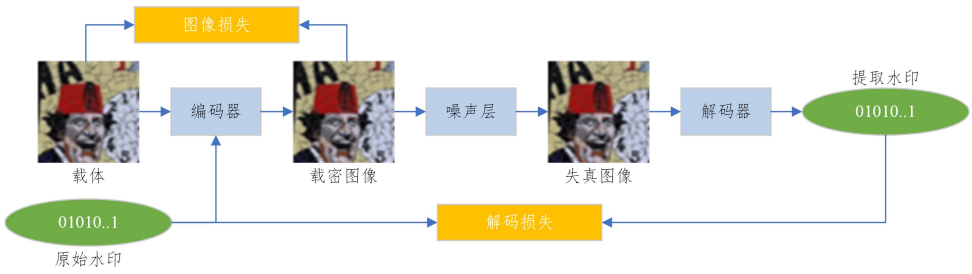


图 1 “编码器-噪声层-解码器”深度水印框架

Fig. 1 “Encoder-Noise Layer-Decoder” deep watermarking framework

其中，编码器部分负责将水印序列以不易被察觉的方式嵌入到原始载体中；噪声层则负责往嵌入水印后的图像中添加噪声，以模拟载体图像在信道传输中造成的失真；解码器负责将水印信号从失真的图像中提取出来。整个框架要求所有部分可导，从而能进行端到端的训练。而在编码阶段，根据是否需要原始图像载体的参与，算法又可以分为自适应编码和

非自适应编码[10]两种。

2.2 抗屏摄攻击水印技术

END 模型的难点在于，算法需要噪声层操作可导，否则就不能对整个框架进行联合训练。对于屏摄等复杂失真的问题，Tancik 等[11]提出使用透视变化、色彩变化、JPEG 压缩等系列操作的叠加，来近似打印拍照带来的失真，从而使得原本不可导的打印拍照过程能通过近似的方式纳入端到端训练过程。在此基础上，为进一步模拟屏摄过程，Jia 等[12]在上述工作的基础上，将拍摄过程引入的投影变换和光线变换通过数学模型近似并以可导的操作加入训练过程。Wengrowski 等[13]则设计了一个网络，用于模拟屏摄失真，为此，他们生成了大量的“屏幕-手机”拍摄的数据用于训练模拟网络，最终将模拟网络加入到噪声层中进行训练，以达到良好的屏摄鲁棒性。但这样的方法有两个缺点：1）生成数据集工程量较大，需耗费大量的人力和财力；2）使用这种方式生成的网络普适性较差，容易出现过拟合的现象，仅对训练集中的样本有较好的表现。

除基于深度学习的方案以外，Cheng 等[14]将消息嵌入摩尔纹以实现屏幕内容溯源，并根据摩尔纹定位到嵌入区域实现水印提取，解码出溯源信息。Guglemann 等[15]基于亮度设计了亮度模板，将模板叠加在屏幕上来实现水印的嵌入和提取。最近，一种用于跨设备通信的图像码“TERA Code”[16]被提出。该方法利用高刷新率屏幕交替显示两帧互补的水印图像，以抵消视觉失真效果，在跨设备通信领域是良好的尝试。

虽然针对屏摄攻击的水印研究已取得一定成效，但由于屏摄过程的复杂多变性，现有方案距离实际应用仍存在距离。如何解决屏摄过程中相对于传统电子信道而言没有的镜头失真、光照失真、摩尔纹失真等难以模拟的干扰带来的影响，仍然是目前针对屏摄鲁棒水印算法的重点研究内容。

3 屏摄鲁棒 DCT 域深度水印方法

本文设计了一种 DCT 域深度水印方法。该方法通过学习修改 DCT 的系数来提高水印嵌入的透明性，通过实验对比测试验证了方法的有效性；此外，本文方法还对摄屏过程中存在的干扰进行模拟，利用数据增强方案对水印网络中的解码器进行增强训练，从而提高了模型对屏摄攻击的鲁棒性。

3.1 模型结构

本文提出的模型由 DCT 层、编码器、噪声层和解码器构成，结构如图 2 所示。

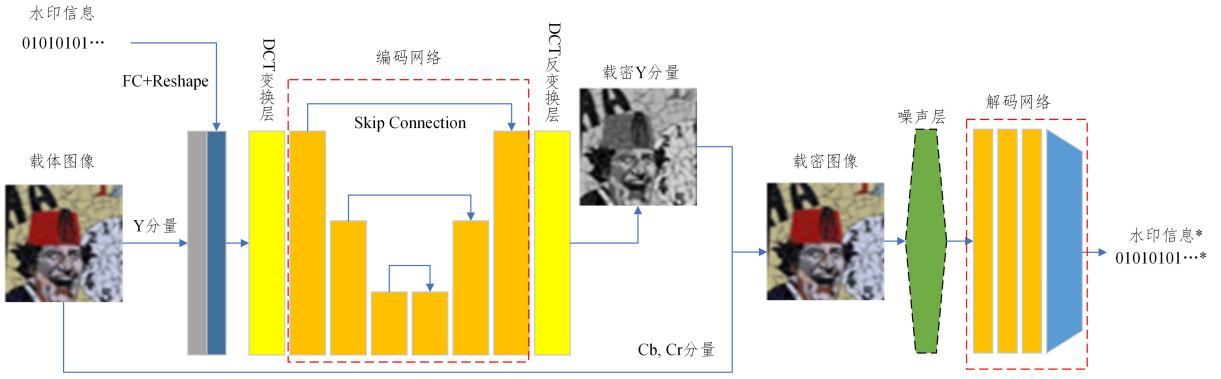


图 2 网络结构图

Fig. 2 Diagram of network structure

3.1.1 DCT 层与编码器

载体图像首先经过 DCT 层，利用 DCT 卷积核将输入的 Y 通道图像变换到 DCT 域。Y 分量的 DCT 系数与整合后的水印信息进行合并生成待编码张量。编码器设计利用 U-Net 网络将输入的 DCT 系数和水印信息进行编码嵌入。编码器进行编码之前先利用 DCT 层对载体图像的 Y 通道进行 DCT 变换，形成大小为 $H\times W\times 1$ 的 DCT 系数张量 $\boldsymbol{C}$。DCT 卷积采用 $H\times W$ 的卷积核，其计算过程如式(1)所示：

$$\begin{aligned} D(\boldsymbol{u},\boldsymbol{v}) &= C(u)C(v)\sum_{0}^{H-1}\sum_{0}^{W-1}\boldsymbol{Y}(i,j)P(i,j)\\ P(i,j) &= \cos\left[\frac{(2i+1)\pi}{2H}u\right]\cos\left[\frac{(2j+1)\pi}{2W}v\right]\\ C(u) &= \begin{cases}\sqrt{\dfrac{1}{H}}, & u=0\\ \sqrt{\dfrac{2}{H}}, & u\neq 0\end{cases} \end{aligned} \tag{1}$$

其中，$\boldsymbol{Y}(i,j)$代表 Y 通道(i,j)位置上的像素值，$D(\boldsymbol{u},\boldsymbol{v})$代表 DCT 系数矩阵$(\boldsymbol{u},\boldsymbol{v})$位置上的系数结果。

秘密信息为大小为 L 的随机信息，秘密信息经过全连接层并上采样成和 Y 通道图片一样大小的张量 $\boldsymbol{S}$，大小为 $H\times W\times 1$。将 DCT 系数 $\boldsymbol{C}$ 和秘密信息 $\boldsymbol{S}$ 进行拼接，形成待编码张量 $\boldsymbol{T}$。张量 $\boldsymbol{T}$ 的大小为 $H\times W\times 2$。过程如图 3 所示。

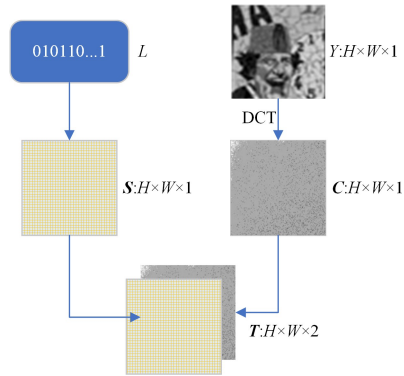


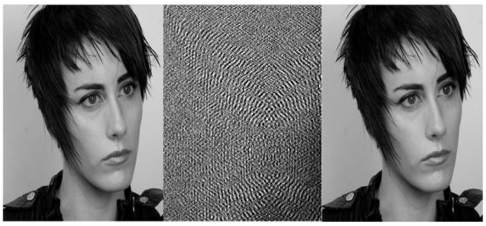
图 3 生成待编码张量示意图

Fig. 3 Example diagram of tensors to be encoded

编码器采用与 StegaStamp[11]相同的 U-Net 网络[17]。将

DCT层的输出张量作为输入，张量大小为 $H\times W\times 2$。编码过程中，进行了4次卷积下采样操作和4次上采样，利用skip connection的操作，将U型网络上采样过程中水平对应的层进行合并以复用前期特征，辅助上采样学习生成嵌入水印后的系数。网络提取输入张量特征，学习修改Y分量DCT系数以实现水印嵌入。卷积层的卷积核大小为3×3，并且设置滑动步长为2以达到下采样的目的。最后一层卷积层输出大小为 $H\times W\times 1$ 的嵌入水印后的DCT系数 C'。

经过编码层修改后的Y分量DCT系数经由反DCT变换层变成嵌入水印信息后的Y分量图。反DCT层的卷积核由DCT层的卷积核进行转置获得。同样地，反DCT层在训练中不参与更新。经由反DCT层后得到Y分量载密图像 Y'，大小为 $H\times W\times 1$。DCT变换与反变换示意图如图4所示。



注：从左到右分别为原图、DCT变换、反DCT变换。

图4 DCT变换与反变换示意图

Fig. 4 Diagram of DCT transform and inverse transform

3.1.2 解码器

解码器基本由卷积层和全连接层组成，如图5所示，用于提取特征并最终映射到与秘密信息长度 L 一致的01序列中，即解码出水印信息。经过卷积层提取载密图像的特征图，卷积层的卷积核大小均为3×3，在增加特征图通道数的同时采用步长为2的卷积实现特征图下采样，最后经由全连接层提取出长度为 L 的信息。

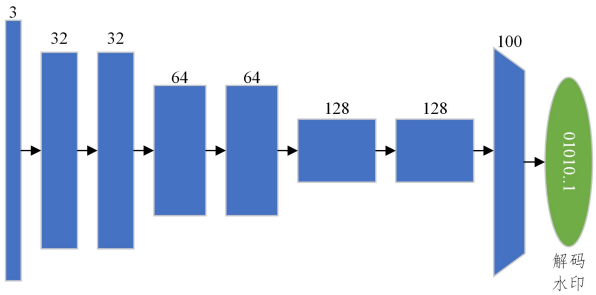


图5 解码器网络结构

Fig. 5 Decoder network structure

3.1.3 噪声层——屏摄模拟

相比传统电子传输而言，摄屏图像存在着最严重的失真，即反光和摩尔纹。本文提出了对应的模拟办法。

视觉上摩尔纹表现为彩色干扰纹，通常表现为波浪状，因此通过对波浪状的彩色波纹进行模拟调整，以达到接近现实中摩尔纹的效果。通过随机函数来选取波纹中心，以确定波纹的起始位置。而对于叠加偏转角的选取，则以弧度为单位，使得绘制点在原始位置偏移一定弧度，从而使得绘制时通过不同“圆形”的叠加形成波纹效果。通过欧氏距离公式计算当前绘制点与波纹中心的距离，并将其作为绘制半径，将纵坐标的高度差作为计算反正弦函数的依据，计算出绘制点距离波纹中心在坐标系中原始的角度大小。通过叠加原始角度和偏转角来形成最终偏转的弧度值，如式(2)所示：

$$D=\sin^{-1}\frac{y_0}{\boldsymbol{R}}+R\times D_t \tag{2}$$

其中，y_0 是当前绘制点与摩尔纹中心的高度差值；R 是绘制半径，是当前绘制点距离摩尔纹中心的欧氏距离；D_t 是偏转角，用于形成绘制偏转，最终形成摩尔纹的波浪效果。

根据偏转弧度，利用正弦和余弦函数计算得出最终需要考虑的像素点的基础坐标值。在基础坐标的基础上，进行上下左右4个方向的偏移，根据每个点对应的颜色强度系数进行4处颜色值的叠加，以形成含有摩尔纹的图像，如式(3)所示：

$$\boldsymbol{I}_m=\boldsymbol{R}_1\times\boldsymbol{I}_1+\boldsymbol{R}_2\times\boldsymbol{I}_2+\boldsymbol{R}_3\times\boldsymbol{I}_3+\boldsymbol{R}_4\times\boldsymbol{I}_4+\boldsymbol{R}_5\times\boldsymbol{I}_5 \tag{3}$$

其中，$\boldsymbol{I}_m$ 是叠加摩尔纹后的绘制点处的像素值；$\boldsymbol{I}_1$，$\boldsymbol{I}_2$，$\boldsymbol{I}_3$，$\boldsymbol{I}_4$，$\boldsymbol{I}_5$ 分别为绘制点自身像素值、向右偏移像素、向左偏移像素、向上偏移像素以及向下偏移像素；$\boldsymbol{R}_1$，$\boldsymbol{R}_2$，$\boldsymbol{R}_3$，$\boldsymbol{R}_4$，$\boldsymbol{R}_5$ 则分别是对应的权重系数，通过权重系数约束绘制点本身和4个偏移量的影响。

在进行叠加时，我们为每个方向偏移的分量分配了一个权重系数，以最终形成颜色混合的结果。同时，由于直接对原图像进行颜色混合的波纹效果会过于突兀，因此计算了颜色叠加后的残差图像，通过残差图权重系数来决定最终叠加的摩尔纹的强弱，如式(4)所示：

$$\boldsymbol{I}=\lambda_m\boldsymbol{I}_m \tag{4}$$

其中，λ_m 是摩尔纹色彩叠加系数，通过控制λ_m的大小来实现摩尔纹色彩混合的强度。

和摩尔纹的模拟一样，反光的模拟需要首先确定反光中心，随后需要确定待绘制的模拟光斑的半径大小。由于现实情况中的反光并不是单纯意义上的白色光斑，而是存在着渐弱效应的反光区域，越靠近边缘，反光效果越不明显，因此根据光强距离衰减原理，在进行光斑的绘制时，光斑强度的叠加需要沿着光斑半径衰减绘制强度以贴近真实情况。根据光照衰减原理，距离光源越远的物体越暗，光强和距离的平方成反比，如式(5)所示：

$$\frac{d_1^2}{d_2^2}=\frac{i_2}{i_1} \tag{5}$$

其中，d_1 和 d_2 是距离光源的距离，i_1 和 i_2 则代表了距离光源 d_1 和 d_2 距离时的光照强度。

根据图像的大小模拟计算光源距屏幕的距离，从而得到反光中心的光照距离，并将其作为基准。随后在反光光斑半径范围内计算待绘制像素点距离反光中心的距离，根据勾股定理计算出其距离模拟光源的距离，利用距离平方反比计算出绘制点光强并进行绘制，如式(6)所示：

$$\boldsymbol{I}_s=\lambda_d\times\boldsymbol{I}\times\frac{D_l^2}{d^2+D_l^2} \tag{6}$$

其中，D_l 是模拟设置的“反光光源”距离图像的距离；d 是绘制反光光斑时当前绘制点距离反光光斑中心的距离，使用欧氏距离作为计算依据；λ_d 是光强衰减系数；I 是基础光照强度，在一定光照强度的基础上叠加比例系数生成，比例系数是

反光半径和图像宽高较小尺度的比值。

不同于实际光源直接近距离照射到屏幕的情况，反光区域的毛边效应会更加严重，并且光照强度的衰减会更加明显。因此，仅仅根据光强距离衰减进行光强计算是不够的。除了根据光强衰减进行计算外，额外衰减系数的设定也很重要。设定光照强度衰减系数λ_d来辅助光强的计算，以使模拟出的反光光斑更贴近现实情况，衰减系数如式(7)所示：

$$\lambda_d = 1 - \frac{d}{\boldsymbol{R}_l} \tag{7}$$

其中，d 为当前绘制点距离反光中心的距离；R_l是反光光斑的绘制半径大小，通过距离差来进一步限定光照强度。

此外，在摩尔纹和反光的基础上叠加一定程度的高斯模糊等处理，用于模拟拍照失焦情况。结合反光、摩尔纹和高斯模糊处理对摄屏图片进行模拟的过程如图 6 所示。

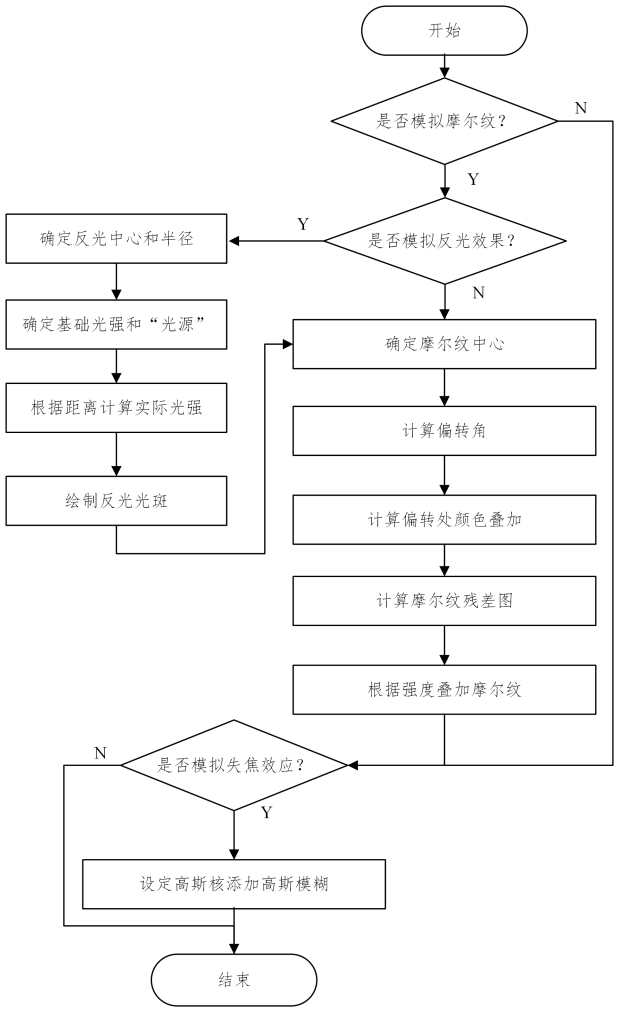


图 6 屏摄攻击模拟过程

Fig. 6 Simulation process of screen shot attack

3.3 训练模式

本文采用两阶段训练的方式。第一阶段只训练 DCT 层、编码器与解码器，第二阶段固定编码器和 DCT 层，使用设计的噪声层对生成的载密图像进行数据增强，再将其送入解码器，对其进行屏摄鲁棒性增强训练。此训练方案不需要噪声层可导。

3.4 损失函数

损失函数分为两部分，分别约束载密图像的生成质量和解码准确度。保证透明性的损失函数如式(8)所示：

$$\boldsymbol{L}_i = \lambda_p \times \frac{1}{H \times W} \| \boldsymbol{L}_p \|_2^2 + \lambda_y \times \frac{1}{H \times W} \| \boldsymbol{I}_c - \boldsymbol{I}_e \|_2^2 \tag{8}$$

其中，$\boldsymbol{L}_p$是载密图像与载体图像的 LPIPS 相似度[18]，H 是图像的高，W 是图像的宽，λ_p和λ_y为权重因子。图像损失的第一部分计算编码前后载体图像和载密图像之间的相似度的均方差。$\boldsymbol{I}_c$是图像 YCrCb 颜色空间的 Y 分量的原图，$\boldsymbol{I}_e$是生成的 Y 分量载密图。计算 Y 分量原图和载密图的均方差作为图像损失的第二部分。

为了保证提取的水印准确无误，解码器需要尽可能使得解码信息和正确的秘密信息接近。由于水印是 01 序列的样式，因此对于解码准确度的损失计算选择交叉熵损失，如式(9)所示：

$$\boldsymbol{L}_m = CrossEntropy(\boldsymbol{M}, \boldsymbol{M}') \tag{9}$$

其中，M 和M'分别为水印信息和解码出的水印信息。

综合编码器和解码器，网络整体的损失函数如式(10)所示：

$$\boldsymbol{L} = \lambda_i \boldsymbol{L}_i + \lambda_m \boldsymbol{L}_m \tag{10}$$

其中，λ_i为编码器的图像 loss 权重，λ_m为解码器的解码准确度权重。

4 实验设计

4.1 数据集

本文采用的实验数据集为 mirflickr25k 数据集[19]，其中 90%的图像用于训练过程，余下的 5%和 5%分别用作验证集和测试集。

4.2 实验准备与相关设置

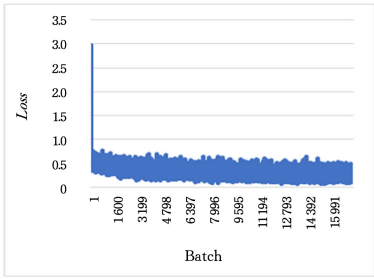
本文采用 GeForce RTX 2060 显卡进行训练，模型框架为 Tensorflow 1.6。输入的图像经过预处理变为 400×400 大小，且数据全部映射到[0，1]之间。秘密信息通过随机函数随机生成长度 $L=100$ 的随机 01 序列，随后再映射到图像大小的张量中，和图像 DCT 系数张量拼接成新的张量送入编码器。模型优化器默认为 Adam[20]，学习率设置为 1×10^{-4}。网络层的初始化方法均采用默认的 Xavier Uniform Initializer。

4.3 评价标准

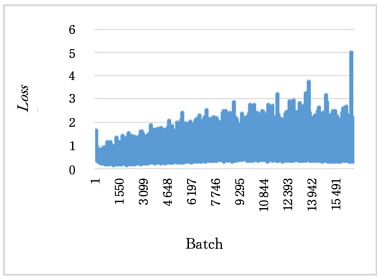
数字水印的性能中，最需要关注的是水印的透明性和鲁棒性。选取峰值信噪比(Peak Signal to Noise Ratio，PSNR)作为对水印算法生成的载密图像质量的客观评价指标，将比特误码率(Bit Error Rate，BER)即错误解码的比特数占总水印信息长度的比例，作为提取水印时的鲁棒性评价指标。

4.4 参数设置

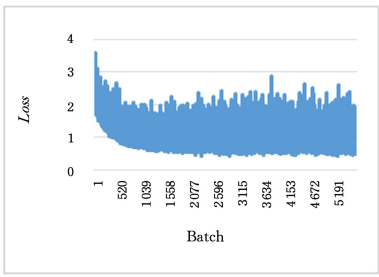
为了更好地让解码器学习到解码特征，在前期训练时，只对编码器损失函数进行优化。设置不同的忽略编码器优化批次 NoI 和图像损失中的 λ_y 进行对比实验，设置为(2000，1.0)，(0，10.0)，(10000，1.0)。对比损失下降情况、图片编码失真情况以及解码准确度来确定最合适的参数和权重，以保证训练的稳定和方法效果最好。不同参数下网络的收敛情况如图 7 所示。图 7 中，横坐标代表当前训练的批次数，纵坐标则代表了当前训练批次下的整体损失大小。



(a)参数为 $N_0I=2000$，$\lambda_y=1.0$ 的损失变化



(b)参数为 $N_0I=0$，$\lambda_y=10.0$ 的损失变化



(c)参数为 $N_0I=10000$，$\lambda_y=1.0$ 的损失变化

注：图 7(a)、图 7(c)均为编码器网络开始更新时的损失变化。

图 7 不同训练条件下的损失变化曲线

Fig. 7 Loss curve under different training conditions

通过对比分析可知，当忽略编码器网络更新的批次太多时，后期训练过程中进行编码器和解码器的整体训练会出现较大的波动。当一开始就进行编码器和解码器的联合训练时，解码器难以在一开始就从隐藏得很好的载密图像中学习到正确的水印知识，因此 loss 下降缓慢且存在一定程度上的波动。而当 NoI 和 λ_y 分别设置为 2000 和 1.0 时，loss 迅速下降，虽然存在波动，但网络整体以平稳的趋势收敛。图像生成质量和解码准确度情况如表 1 所列，根据结果，将实验参数设置为忽略图片质量优化 2000 批次，Y 分量图像损失权重设置为 1。

表 1 不同权重参数下的 PSNR 和 BER

Table 1 PSNR and BER with different weight parameters

(NoI,λ_y)	PSNR	BER
(2000,1.0)	40.16	1.89
(0,10.0)	41.34	4.12
(10000,1.0)	36.73	1.36

在进行摄屏攻击模拟时，使用随机函数对载密图像随机进行摩尔纹、反光干扰。反光的概率设置为 50%，摩尔纹的出现概率设置为 62.5%，高斯模糊的概率为 25%，高斯核大小为 3×3。光照和摩尔纹中心区域设置在图像中间偏中上区域。摩尔纹和反光的模拟的真实度也影响着增强后解码器网络的泛化性。反光光照强度的大小、摩尔纹的偏转角度大小、摩尔纹色彩混合强度等都是需要考虑的因素。本文设置不同的光强大小、偏转角大小、混合强度进行实验。

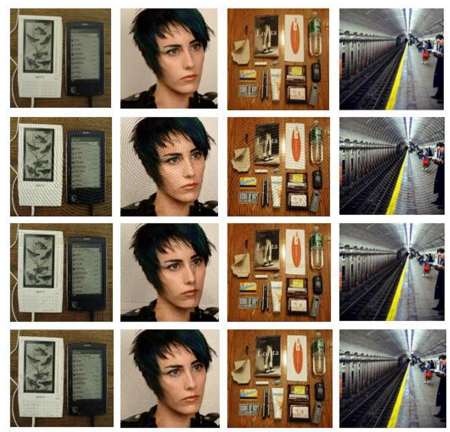
不同光照强度反光模拟图如图 8 所示。由图可知，光照强度设置为 50 时，反光光斑像素值过低，光斑不够明显，不符合反光光斑的局部高亮度特点；光照强度设置为 800 时，光照中心至边缘区域的过渡不明显，形成了很突兀的光斑，和现实中反光光斑存在一定的差距；光照强度为 300 时，光斑亮度合适，中心至边缘过渡自然，较为贴近真实反光情况。因此，将强度基准值设置为 300 比较合适。同时，光斑区域的半径大小也要适中。不同于近距离的直射光源，反光所造成的光斑往往不会特别大，选取图像宽高尺寸中较小的一个尺度，让光斑半径在该尺度的[1/8,1/6]区间随机波动是比较合理的。

注：从上至下依次为原图、光强 50、光强 300、光强 800。

图 8 不同光照强度反光模拟图

Fig. 8 Reflection simulations under different light intensity

摩尔纹的偏转角大小影响着叠加在图像上的干扰纹路波纹的大小和形状分布。偏转角过大时，形成的摩尔纹的波纹弧度过小并且分布集中，视觉上更像是椒盐噪声的效果，与现实中摩尔纹情况不符。当偏转角过小时，生成的摩尔纹由于偏转不够导致波纹效果不明显，摩尔纹呈规律树轮状，波纹数量过少，也不符合实际情况。当设置偏转角大小为 π 时，绘制的摩尔纹波纹数量、弧度较为正常。不仅有波纹状的摩尔纹分布，也有弧度不明显的竖条状摩尔纹分布。因此，设置摩尔纹的偏转角大小为 π 时，绘制的摩尔纹能较好地模拟现实中摩尔纹的样式，该情况下的摩尔纹最接近实际情况。不同偏转角的效果如图 9 所示。

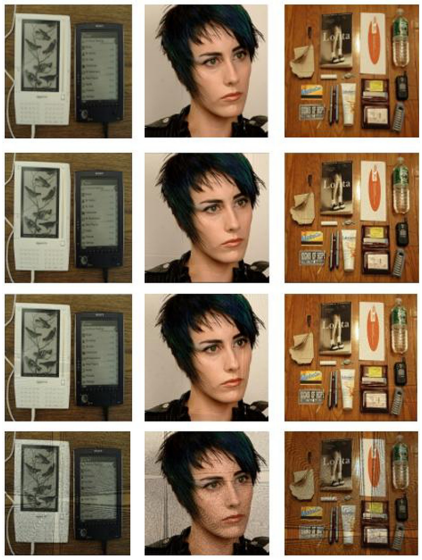


注：从上至下分别为原图、π/6、π、7π/2。

图 9 不同偏转角摩尔纹模拟图

Fig. 9 Moire pattern simulation with different deflection angles

摩尔纹的色彩混合强度由 4 个偏移分量组合而成，上下左右 4 个分量的权重系数均在[0.1，0.8]区间随机进行色彩的混合，以贴近现实情况中摩尔纹的色彩失真效果。4 个分量总体的混合强度系数也对摩尔纹的模拟起到了决定性作用。混合强度过大时模拟摩尔纹的色彩失真颗粒感明显，摩尔纹绘制过重，不符合现实情况。混合强度过小时波纹偏向于灰度化，也不符合要求。因此设置总体权重为 0.8，此时波纹的色彩混合强度适中，绘制的摩尔纹的色彩失真较为贴近真实摄屏情况下存在的摩尔纹。不同色彩混合强度效果如图 10 所示。



注：从上至下分别为原图、0.05、0.8、3。

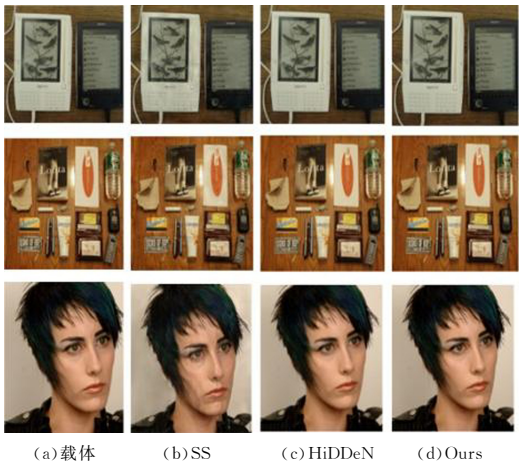
图 10 不同色彩混合强度摩尔纹模拟图

Fig. 10 Moire pattern simulation with different color mixing intensities

5 实验及结果分析

5.1 载密图像视觉质量对比实验

首先，将提出的方法和目前比较经典的深度水印 StegaStamp(SS)[11]和 HiDDeN[4]进行视觉质量对比。图 11 给出了 3 种水印方法生成的载密图像示例。



(a)载体 (b)SS (c)HiDDeN (d)Ours

图 11 不同水印方法的载密图像示例

Fig. 11 Examples of container images of different watermarking methods

可以发现，本章设计的方法能较好地保存载体的细节和颜色特征，在视觉观感上的失真也明显降低。此外，我们使用 PSNR 指标对生成的载密图像进行图像质量的评价，结果如表 2 所列。可以明显地看到，主观视觉较差的 HiDDeN 生成的载密图像的 PSNR 值也较低，StegaStamp 在 PSNR 值上和本章提出的方法比较接近。本章提出的方法的 PSNR 指标达到了 40.1，优于其他方法。

表 2 不同方法生成的载密图像的 PSNR

Table 2 PSNR of container images generated by different methods

Methods	PSNR
SS[9]	38.2
HiDDeN[2]	34.2
Ours	40.1

5.2 鲁棒性对比实验

不同的噪声和噪声强度对于水印解码准确率的影响存在着差异。首先针对电子信道噪声，将增强训练后的解码器和 HiDDeN、SS 以及在抗摄屏领域有较好表现的 ARIW[9]方法进行对比，结果如表 3 所列。裁剪攻击随机去除∂的载密图像边界，并在解码前缩放为原始图像大小。JPEG 压缩在[50，100]中随机选择质量因子 q，高斯噪声和椒盐噪声在[0，0.2]中随机选择噪声系数 σ。

表 3 不同方法在不同噪声攻击下的误码率

Table 3 Bit error rates of different methods under different noise attacks

methods	裁剪		JPEG	高斯噪声	椒盐噪声
	$\partial=0.1$	$\partial=0.4$	$q\in[50,100]$	$\sigma\in[0,0.2]$	$\sigma\in[0,0.2]$
ARIW[9]	0.19	2.45	8.62	4.06	6.23
HiDDeN[4]	1.87	7.23	17.46	27.89	38.45
SS[11]	2.48	16.62	6.12	1.89	5.14
Ours	3.46	18.23	5.64	1.18	5.37

由于本文采取在载体图像的 Y 分量 DCT 系数上进行水印嵌入，频域系数的修改对于常见电子信道的噪声有一定的抵抗能力，因此和使用相同噪声层结构的 SS 相比，本文方案性能与其接近的同时在抗 JPEG 压缩和高斯模糊上性能略优。但除了由于水印位置信息丢失导致的解码器抗裁剪能力劣于对比方法，增强训练后的解码器在椒盐噪声干扰下的鲁棒性的表现也不如对比模型，这是由于椒盐噪声大多是随机干扰，并且表现为突兀的颗粒状，对于修改频域系数来嵌入水印的方法而言干扰略大。但总体而言，增强训练后的解码器在常见噪声干扰下的表现较好。

对载密图像进行真实场景下的随机摄屏，比较不同方法的解码准确率。不同方法的检测误码率如表 4 所列。

表 4 不同方法在真实屏摄场景下的检测误码率

Table 4 Bit error rates of different watermarking methods in real world screen shoot attack scenario

methods	BER
ARIW[9]	5.23
HiDDeN[4]	21.36
SS[11]	4.58
Ours	2.84

由表 4 可知，HiDDeN 在摄屏情况下的误码率较高。摄屏图像存在电子信道噪声和摄屏攻击等多种干扰，大大影响了 HiDDeN 方法的解码准确率。基于 ARIW 的水印方法和 SS 在摄屏情况下的误码率在 6% 以下，效果优于 HiDDeN，但其在存在摩尔纹或者存在反光情况的摄屏图像下的解码效果并不理想。本文提出的模型在摩尔纹和反光失真模拟增强下取得了最优的效果。

我们进一步测试了不同角度和距离的拍照时的解码准确度，如表 5 所列。表 5 中，LR 代表从屏幕左右偏转角度拍摄，UD 代表从屏幕上下偏转角度进行拍摄。在拍摄角度较小时，解码器仍能较为准确地解码出水印信息，误码率基本在 8% 以下。当摄屏镜头与显示器角度过大时，解码准确度下降明显。特别地，当拍摄存在仰俯角时，误码情况比较严重，但也保持在 20% 以下。在不同情况下，本文方案均优于其他方法。

表 5 不同拍摄距离和拍摄角度下的误码率对比结果

Table 5 Comparison results under different shooting distances and angles, the BERs are reported

距离/cm 角度/(°)	20 0	30 0	60 0	30 20(LR)	30 20(UD)	30 50(LR)	30 50(UD)
ARIW[7]	2.41	2.71	7.98	5.12	7.56	16.12	22.00
SS[9]	2.30	2.93	7.23	4.56	7.32	15.62	20.44
Ours	2.27	2.84	7.19	3.36	6.67	13.89	18.15

随着移动设备的快速发展，不同型号的移动设备层出不穷，不同移动设备的拍照情况以及可能存在的摄屏攻击干扰程度都有所不同。因此，我们使用不同型号移动设备进行摄屏拍照并解码，分析在不同移动设备拍照情况下本文方法对于摄屏攻击的鲁棒性，情况如表 6 所列。

表 6 不同型号移动设备摄屏下的水印误码率

Table 6 Bit error rates of different watermarking methods under different types of mobile devices

	Iphone 8	MI K40	1^+	P40
ARIW[7]	3.13	4.56	3.50	3.80
SS[9]	3.00	4.5	3.20	3.90
Ours	2.84	4.27	3.13	3.71

对于不同型号的手机摄屏测试，基于拍摄距离 30 cm、零拍摄角度进行。由表 6 可知，iphone 8 和 1^+ 手机拍摄的水印图片解码时误码率低于小米 K40 和华为 P40 手机。这是由于后两者的系统相机往往存在着美颜、滤波处理。整体而言，对于不同型号手机的摄屏图片，解码器基本能以较小的误码率解码出水印信息。同时，本文方法针对不同设备间的鲁棒泛化性也优于其他方案。

结束语 本文针对目前深度水印模型透明性较差的问题，提出了 DCT 域的深度水印方案。具体来讲，本文将秘密信息嵌入载体图像 Y 通道的 DCT 系数中，DCT 域系数的修改使得嵌入水印对图像的影响分散到图像全部区域，减少了可能会出现在轮廓等特征中的明显视觉失真。通过训练自学习修改过程，兼顾了水印嵌入的透明性和鲁棒性。此外，本文分析了屏摄失真的特性，提出了摩尔纹、反光噪声模拟方案。在 DCT 域深度水印方案的基础上，利用数据增强的方式将摩尔纹失真和反光噪声加入水印模型的端到端训练过程，保证了水印模型抵抗屏摄失真的性能。对比实验结果表明，本文模型对传统失真和屏摄失真的鲁棒性均优于同类方法。

参 考 文 献

[1] LIU S W, DU Q Z, LONG H, et al. Robust Audio Watermark Based on QR Code[J]. Journal of Chinese Computer Systems, 2022, 43(7): 1535-1540.

[2] LI X Y, ZHOU X B, LIU Z. High-capacity and robust image watermarking algorithm[J]. Journal of Jilin University (Engineering and Technology Edition), 2022, 52(1): 174-179.

[3] FANG H, ZHANG W, ZHOU H, et al. Screen-shooting resilient watermarking[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(6): 1403-1418.

[4] ZHU J, KAPLAN R, JOHNSON J, et al. Hidden: Hiding data with deep networks[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 657-672.

[5] JOHNSON N F, JAJODIA S. Exploring steganography: Seeing the unseen[J]. Computer, 1998, 31(2): 26-34.

[6] SHEISI H, MESGARIAN J, RAHMANI M. Stegano-graphy: Dct coefficient replacement method and compare with JSteg algorithm[J]. International Journal of Computer and Electrical Engineering, 2012, 4(4): 458-462.

[7] KUNDUR D, HATZINAKOS D. A robust digital image watermarking method using wavelet-based fusion[C]// Proceedings of International Conference on Image Processing. IEEE, 1997, 1: 544-547.

[8] FANG H, CHEN D, HUANG Q, et al. Deep template-based watermarking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(4): 1436-1451.

[9] ZHONG X, HUANG P C, MASTORAKIS S, et al. An automated and robust image watermarking scheme based on deep neural networks[J]. IEEE Transactions on Multimedia, 2020, 23: 1951-1961.

[10] ZHANG C, BENZ P, KARJAUV A, et al. Udh: Universal deep hiding for steganography, watermarking, and light field messaging[J]. Advances in Neural Information Processing Systems, 2020, 33: 10223-10234.

[11] TANCIK M, MILDENHALL B, NG R. Stegastamp: Invisible hyperlinks in physical photo-graphs[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2117-2126.

[12] JIA J, GAO Z, CHEN K, et al. RIHOOP: robust invisible hyperlinks in offline and online photo-graphs[J]. IEEE Transactions on Cybernetics, 2020: 7094-7106.

[13] WENGROWSKI E, DANA K. Light field messaging with deep photographic steganography[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1515-1524.

[14] CHENG Y，JI X，WANG L，et al.｛mID｝：Tracing screen photos via｛Moiré｝patterns[C]// 30th USENIX Security Symposium (USENIX Security 21). 2021：2969-2986.

[15] GUGELMANN D，SOMMER D，LENDERS V，et al. Screen watermarking for data theft investigation and attribution[C]// 2018 10th International Conference on Cyber Conflict(CyCon). IEEE，2018：391-408.

[16] FANG H，CHEN D，WANG F，et al. TERA：Screen-to-Camera Image Code with Transparency，Efficiency，Robustness and Adaptability[J]. IEEE Transactions on Multimedia，2021，24：955-967.

[17] RONNEBERGER O，FISCHER P，BROX T. Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention. 2015：234-241.

[18] ZHANG R，ISOLA P，EFROS A A，et al. The unreason-able effectiveness of deep features as aperceptual metric[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018：586-595.

[19] HUISKES M J，LEW M S. The mir flickr retrieval evaluation [C]// Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval. 2008：39-43.

[20] KINGMA D P，BA J. Adam：A method for stochastic optimization[J]. arXiv：1412. 6980，2014.

HUANG Changxi，bornin 1986，Ph. D. His main research interests include information hiding and digital watermarking.

LING Hefei，born in 1976，Ph.D，professor，Ph.D supervisor，is a senior member of CCF(No. 05610S). His main research interests include computer vision and so on.

（责任编辑：喻藜）

一项针对高科技企业员工的调查显示：创新能力并非随年龄增长而降低

中青报·中青网记者　李　新

在高科技企业中，除了专业技能这样的“硬能力”，还有哪些“软能力”影响员工的发展？不同业绩水平的员工在“软能力”上有什么差别？不同性别、不同年龄、不同学历员工在“软能力”上是否有区别……一项有关高科技企业一线员工软能力的调查回答了这些问题。1 月 16 日，《企业一线员工软能力调研报告》发布，很多调查结果与之前普遍的社会认知有所不同，例如创新能力并非随年龄增长而降低，并非学历越高软能力越强。

这项调查是由中国计算机学会和北京中关村人才协会共同组织的。历时近两年完成了对 20 多家大中型高科技企业中 63 名一线经理的访谈，评价了 450 多名 35 岁以下一线员工的软能力。软能力是社会心理学术语，它们是人格特质、社交礼仪、语言沟通能力、个人习惯、品德和乐观态度的体现。员工软能力的不足，对企业运行效率影响颇大，是企业整体能力提升的障碍。曾有企业高管对员工软能力不满意——“新员工入职培训要重新做幼儿园和小学的教育”。

本次调研的 13 项软能力，包括高效执行、积极进取、责任意识、风险意识，言语理解、逻辑思维、批判性思维、学习创新，情绪调控、环境适应、沟通协调、团队合作、同理心。

调研发现，批判性思维、风险意识、学习创新是一线经理对员工评价最低的 3 项软能力，同理心、情绪调控、沟通协同 3 项的评价也不高。同时，被评价员工年龄越小，上述几项软能力的评分越低，31～35 岁年龄段人员的软能力明显高于 21～25 岁、26～30 岁两个年龄段。

业绩优秀员工软能力高于业绩良好和一般人员；男员工和女员工的软能力得分总体接近，女性的责任意识、沟通协调的表现优于男性，男性的积极进取、言语理解、逻辑思维、情绪调控和风险意识表现优于女性。

在学历对比上，此次调研的结果出乎调查者意料。硕士学位员工在高效执行、言语理解、学习创新和批判性思维等几项软能力的得分竟落后于本科生。博士生优于其他学历者。

1 月 16 日，报告发布的同时，举办了一场“软能力培养研讨会”，调查者提出企业“搞硬科技需要提升软能力”。同时，基于调查结果，对企业用人“学历军备竞赛”提出了不同的看法，认为应更关注员工各方面综合能力，不要唯学历论。企业在技术培训和技能培养上应承担更多责任；大学则应利用学生在校时间长的特点，在软能力等基础培养上下更大力气；企业与大学在人才培养上要避免错位，应各自回归本位。

此外，调研报告也提出教师和家长应以身教为先，言传为次，在潜移默化中完成软能力的培养任务；软能力培养应多角度、多方面、多批次落实，不能寄希望于某一项或几项特定活动；软能力培养应从幼儿抓起。

据 CCF 微信公众号