《隐写安全性增强与分布保持隐写研究》整理

原创

岁月漫长_ 已于 2022-03-21 17:11:52 修改 2639 收藏

分类专栏： 图像隐写 文章标签： 计算机视觉 人工智能

于 2022-03-03 20:12:00 首次发布

本文链接：https://blog.csdn.net/qq_40859587/article/details/123185954

绪论

信息论下的隐写安全性：一个隐写系统，载体集合的分布为P，隐写后得到的载密集合分布为P，隐写安全性定义为两个分布之间的**相对熵（KL散度）**（难以计算，《隐写学技术与原理》上说用MMD最大平均偏差来替代 P12）

$$D(P_c||P_s)=\sum_{x\in\mathbf{X}}P_c(x)\log_2\frac{P_c(x)}{P_s(x)}. \tag{1.1}$$

如果 $D(P_c|P_s)\leq\epsilon$, 则称隐写方法对于被动攻击者是 ϵ 安全的。当 $D(P_c||P_s)=0$, 则称隐写方法是绝对安全的。

空域图像隐写

1. 非自适应隐写：LSB±1，改进思路提高嵌入效率，增加单位修改承载的消息容量，减少修改量，如MMD，ZZW编码。
2. 自适应隐写：有了STC编码后，定义合理的失真函数（图像纹理复杂程度）成为主要研究内容。
    1. 空域图像中，**HUGO**算法（知乎上一个论文解析）根据载体图像上修改的隐写分析特征相对于原始图像特征的偏差来定义失真，偏差小的代价小。
    2. 但是HUGO大量修改点集中在**图像光滑边缘**区域，对此，**WOW**和**UNWIARD**利用**多方向小波滤波残差**定义失真，在多个方向上都难以被预测的像素，即滤波残差大的像素赋予较小的失真。
    3. HILL隐写失真函数，Li等人讨论了嵌入失真的分布对抗检测性能的影响，提出三个隐写失真定义原则：**纹理复杂度优先**原则（HUGO，UNIWARD）、**失真扩散**原则（HILL）和**修改聚集**原则，并根据原则提出了隐写失真函数**HILL**
    4. Sedighi等人用多元高斯分布对载体图像像素建模，根据最小化该模型下载体载密图像分布的KL散度的目标，设计失真函数，提出了MVG隐写算法。在MVG的基础上，根据假设检验原理，在最小化似然比检验子指导下，推测像素分布模型并对其参数进行估计，设计了**MiPOD**隐写失真定义算法

JPEG图像（社交网络常见）隐写

对量化后的**DCT系数**进行隐写。一般DCT系数频率越高，修改失真越大，用量化步长来区分。UED使用非0系数数量定义纹理复杂度。

研究发展

隐写分析：从SPAM的686维到**SRM**的34671维(30个高通滤波器，论文有详细步骤)，Xunet用CNN做空域和JPEG隐写分析，SRnet 超过了传统机器学习隐写特征。
深度学习中GAN可用生成器嵌入信息，判别器提取信息；或用GAN学习失真函数定义。

最小化失真隐写

最优概率分布于失真的对应关系遵循指数分布（Gibbs分布），STC编码可逼近最优性能。

隐写编码

矩阵编码为基础，利用线性分组码的奇偶校验矩阵完成消息的嵌入和提取。

失真定义原则

1.纹理复杂度优先

当前基于纹理复杂度优先原则的代表算法有HUGO（邻域预测残差刻画图像纹理复杂程度），UNIWARD（小波滤波器滤波残差），AACbased（音频）

2.失真扩散

考虑当前像素修改，那么周边像素也适合修改，从而使修改点聚集，提升隐写安全性。一般通过**低通滤波器**对失真进行平滑滤波，达到具有小失真的元素周边的元素也倾向于具有较小的失真，具有大失真的元素周边的元素也倾向于具有较大的失真的效果。

在 HILL 中，使用高通滤波器 $\mathbf{H}$ 的滤波残差来刻画图像的纹理复杂程度，用低通滤波器 $\mathbf{L}_1, \mathbf{L}_2$ 来实现失真扩散。

$$\rho = \frac{1}{|\mathbf{X} \otimes \mathbf{H}| \otimes \mathbf{L}_1} \otimes \mathbf{L}_2 \tag{2.31}$$

其中 $\otimes$ 为卷积操作，高通滤波器 $\mathbf{H}$ 为如下 3×3 的形式:

$$\mathbf{H} = \begin{bmatrix} -1 & 2 & -1 \\ 2 & -4 & 2 \\ -1 & 2 & -1 \end{bmatrix}$$

$\mathbf{L}_1$ 和 $\mathbf{L}_2$ 则分别为 3×3 和 15×15 的均值滤波器。文献中[23] 的实验表明了失真扩散原则提高了隐写的安全性能。

3.方向一致性原则（针对非加性隐写模型）

让邻域元素的修改方向趋向一致，尽可能的保持邻域相关性。

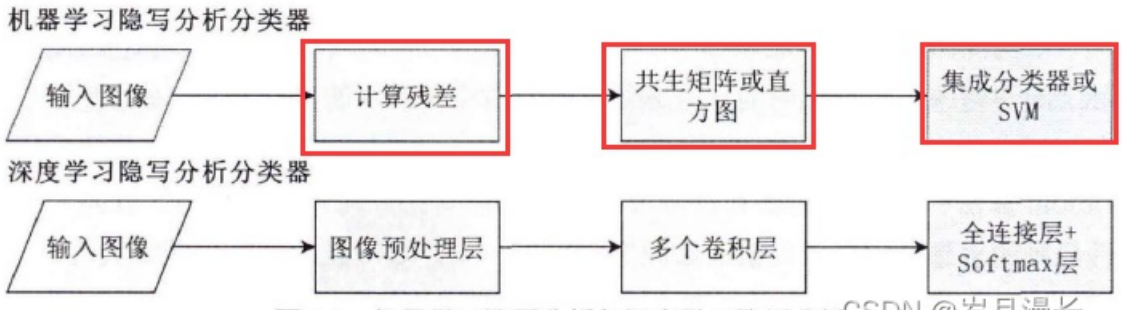

图 2.2 机器学习隐写分析与深度学习隐写分析

第3章 图像微尺度隐写失真模型

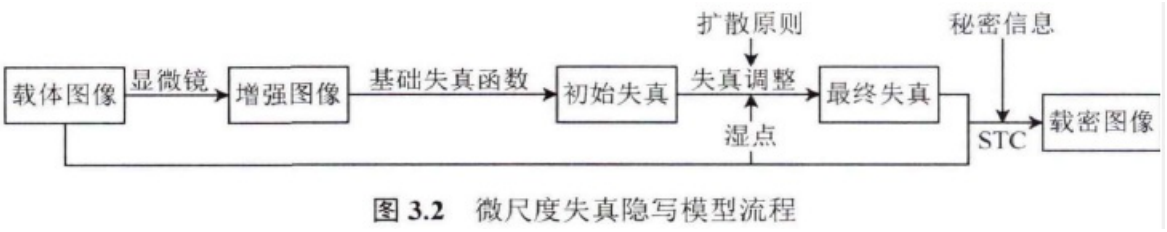

图 3.2 微尺度失真隐写模型流程

作者采用**锐化滤波器USM**对高频区域进行增强：用高通滤波器进行滤波得到高频分量，然后乘上锐化系数，与原始图像叠加得到增强图像。在增强图像上进行失真定义。

隐写的**边信息**：获取JPEG图像压缩之前的空域图像。

实验部分

BOSS图像库，嵌入率分别为

0.1,0.2,0.3,0.4,0.5　比特每像素。SRM和maxSRM作为隐写分析特征，基于Fisher线性分类器的集成分类器用于

深度生成模型（生成新数据）

现有的具有代表性的深度学习生成方法有变分自编码器，对抗生成网络，自回归和基于流的生成模型。其中变分自编码器，对抗生成网络和基于流的生成模型属于隐式生成模型，自回归生成模型属于显式生成模型。

未来展望

利用深度学习模型的图像增强技术，深入挖掘图像像素间的相关性，更准确的定义隐写修改的失真。此外，隐写失真定义和隐写分析也是一个零和博弈过程，利用**对抗生成网络**，由机器设计隐写失真函数也是一个重要的研究方向。