一种端到端的对抗生成式视频数字水印算法

作者：崔凯元，申静，李叶凡，王晗，王忠芝 时间：2021年

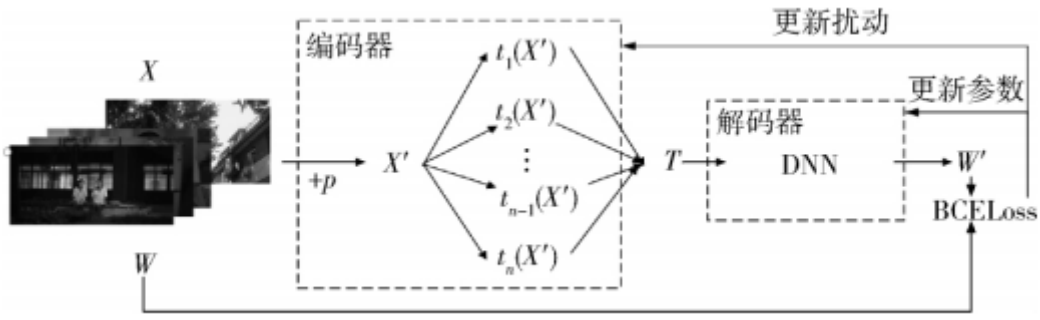
期刊：中国科技论文 第16卷第7期

背景：

- 因为计算机技术和多媒体系统的飞速发展使以数字形式出现的视频内容迅速增长，数字视频成为了网络多媒体、监控系统、远程会议、视频网站等应用中不可或缺的组成部分。数组视频的传播，导致视频数据容易被非法复制、篡改，严重侵犯了视频原作者的版权
- 因此有研究者通过在视频中加入不可见的盲水印，用于版权的保护。但无论是图像水印还是视频水印，传统水印算法对于几何形变的攻击抵抗能力较弱。这是由于即使是微小的变换，也会导致全部像素值的改变，最终无法有效的提取出水印。因此，该文提出一种端到端的对抗生成式视频数字水印算法，为关键帧添加扰动生成对抗性样本，并模拟对抗性样本的各种变换，以抵抗几何攻击。
- 区别于图像水印，视频水印还会受到时间同步攻击，例如帧插入，帧删除，帧平均和帧重组。现有的算法大部分基于固定的一组视频帧进行嵌入水印操作，因此提取水印时需要检测出准确的帧位置。然而在收到时间同步攻击之后，很难从视频中准确定位出这些包含水印的和视频帧
- 视频水印的提取分为盲检测和明检测2种，在提取水印的过程中，不需要对原始视频数据作参考比对的叫作盲检测，对应的水印算法为盲水印算法；反之为明检测和明水印算法。实际应用种，前者应用更广

1.基于解码器的视频数字盲水印算法：

- 算法主要包括有编码器和解码器2部分
    - 编码器根据Goodfellow等提出的快速递度符号法(FGSM)为关键帧X添加扰动p，生成对抗性样本X'，表示为X' = X + ε * p，其中ε为扰动强度，取值在0，1之间，其越大，扰动越明显，在降低模型精度方面越有效，但同时原始图像的改变也越容易被人眼察觉。
    - 解码器D由1个深度神经网络构成，将对抗性样本解码为与水印W长度相同的序列W'，表示为W' = D（T，θ），其中θ为模型参数
- 编码器与解码器采用对抗的方式进行优化。编码器通过加入扰动p，生成让解码器解码错误的对抗性样本；解码器将对抗性样本解码成功，即W' = W。通过不断迭代对抗训练的过程，学习可得可抵抗几何攻击的嵌入水印。



2.水印的嵌入和提取：

- 首先读取视频，抽取视频的关键帧，记录关键帧在原视频中的位置，对于每个关键帧，都进行图像归一化操作，随机初始化解码器的参数和扰动p的大小，开始训练网络，根据原始的关键帧和扰动生成对抗性样本X'，将X'作为网络的输入。

- 提取水印时，首先读取视频，抽取视频的关键帧，从第一个关键帧开始，每一帧重复以下的操 作，即将关键帧作为解码器的输入，解码得到与水印长度相等的一维数组，直到提取出来的序列前32位与嵌入的水印标志位相匹配，提取水印过程结束。

3.结论与总结

- 该文通过实验分析验证证明了这种端到端的对抗生成式视频数字水印算法在对抗JPEG压缩、运动模糊、椒盐噪声、缩放、平移、裁剪和时间同步攻击方面具有较好的鲁棒性。未来工作将在此基础上着重考虑针对旋转等攻击的鲁棒性。
- 之前在网上偶然也看到过相关的视频介绍对抗生成网络与数字水印融合的算法，这类算法的大致思路都是通过生成器和监督器对水印的生成以及检测进行不断地迭代，在互相对抗中最后生成的水印有更好的鲁棒性已经透明性，在该文中，更是提出使用这类算法能够好的防止一些几何攻击，能更加完善水印的功能性和强壮性。