笔记：两篇使用深度学习进行隐写的文章

原创

g28_gerwulf 于 2019-11-30 13:29:11 发布 1163 收藏 10
分类专栏： 机器学习 CV 文章标签： 深度学习 CV

本文链接：https://blog.csdn.net/g28_gwf/article/details/103323602


机器学习 同时被 2 个专栏收录
7 篇文章 0 订阅
订阅专栏


CV
4 篇文章 0 订阅
订阅专栏

Hiding Images in Plain Sight: Deep Steganography（源码）
目标是将N*N*RGB的像素图（Secret Image）隐藏到另一幅N*N*RGB的图片（cover image）中，并使得cover image的扭曲程度最小。允许恢复过来的图片有一定的质量衰减。

流程：
1.使用Prep网络预处理Secret image的扭曲程度最小。允许恢复过来的图片有一定的质量衰减。
2.使用Hiding网络将处理后的Secret image隐藏到Cover Image中
3.使用Reveal NetWork将Secret image恢复出来
Prep网络的两个作用：
1.防止Secret Image小于Cover Image，Prep网络逐渐增加Secret Image的尺寸，使得两者达到相同大小
2.转化有色像素成为更容易被编码的特征，如纹理和边缘
Hiding网络
将Prep网络的输出和Cover Image作为输入并生成Container Image。
Reveal网络
解码器网络将Cover Image移除并显示出Secret Image。

原理
文章的基本的重点在于具体地展示了将大量的信息编码到一张图片且图片只有很少差异的是有可能的。然而，仍没有具体的尝试去主动地使信息的存在从机器检测下隐藏。尽管我们不能期望能完全隐藏最多有1/2的信息是隐藏消息的一部分的事实，仍然能采取措施使得信息很难被探测到。然而我们必须确定Secret Image被隐藏到了哪里。
RGB图中每个像素包括24bit。如果对Container ImageR通道的所有像素的第一个bit进行细微调整，我们可以衡量这对Container Image本身和对重建过程的影响。
Secret Image的信息是分布在所有颜色的通道中的。
多项研究证明深度网络更适合在手选图片上进行隐写分析。
我们对隐写分析网络训练成为一个二分类器，使用未处理的ImageNet图片作为反例，Container Image作为正例。我们发现隐写分析网络在24*N*N的二元输入层*上表现的很好。每24层代表一个输入图像上的像素。
FCN有能力取得90%-95%的结合分类率*。为了对比，相同的网络去训练寻找LSB（Cover Image上的每个通道的L个LSB被换为Secret Image上的），L分别取1、2、3、4，全部取得了超过99%的结合率。一个额外的测试证明了当一个bit被随机放置在4个位置，检测率仍然超过99%。

讨论和将来的工作
如果攻击者得到大量的Container Image，而且也得到了对应的Cover Image或Secret Image，也许攻击者有能力重构出系统。如果攻击者不能得到大量的ground truth数据，可能会使用启发式或者盲目图片分离方式。
由于Secret Image和Cover Image的信息量的比例，检测率率是很高的。我们展示了一种通过减少残差的方法使得通过残差法恢复隐藏图片内容更加困难。
在隐藏图片之前，将像素使用一种排列方式进行重组（共M种排列方法）。这种方法需要（1）发送key（2）提高系统的编码能力。

Steganographic Generative Adversarial Networks
允许使用DCGAN通过标准隐写算法来生成隐写安全的信息嵌入
在经典GAN的基础上增加了一个由CNN实现的判别器S对图像进行隐写分析
使用的隐写算法是正负一嵌入算法，也就是LSB，嵌入率为0.4，且只在图像的一个通道内嵌入秘密信息。

网络结构
Alice，生成网络，产生逼真的图片包含隐藏信息
输入：密钥（二进制），需要隐写的信息（二进制），生成图片的类别（y），噪声
输出：包含被隐写信息的图片
Bob，解密网络，提取图片中的隐藏信息
输入：图片，密钥，图片的类别（y）
输出：被隐写的信息
Discriminator，辨别器，检测图片是生成的还是真实的
输入：图片，图片的类别（y）
输出：0/1（图片是生成的还是真实的）
使用名人数据集，200000张图片，大小64*64。10%的图片作为测试集
A为训练集，B为测试集，隐写算法Stego(x)。在隐写信息后训练集为A+Stego(A)，测试集为B+Stego(B)，训练8个ep