# 隐写分析论文整理[A Siamese CNN for Image Steganalysis]

原创

岁月漫长_ 已于 2022-03-17 17:23:34 修改 1899 收藏 1

分类专栏： 图像隐写 网络安全 文章标签： cnn 深度学习 机器学习

于 2022-02-19 20:38:47 首次发布



图像隐写 同时被 2 个专栏收录

13 篇文章 2 订阅

订阅专栏

网络安全

12 篇文章 0 订阅

订阅专栏

论文：A Siamese CNN for Image Steganalysis

源码：https://github.com/SiaStg/SiaStegNet

## Siamese网络：

参考链接：https://blog.csdn.net/ybdesire/article/details/84072339

**背景**

人脸识别中的one-shot问题：公司员工进行人脸识别，每个员工只给一张照片（训练集样本少），并且员工会离职、入职（每次变动都要重新训练模型）。有这样的问题存在，就没办法直接训练模型来解决这样的分类问题（是不是同一个人）。

为了解决one-shot问题，训练一个模型来输出给定两张图像的**相似度**，模型学习得到的是similarity函数。Siamese网络（模型）能通过学习得到similarity函数。

**Siamese网络原理**

给出两个图像X1和X2的相似度，输出一个向量（比如一维向量）。

- 若两张图为同一个人，两个模型输出的一维向量的欧式距离较小，否则较大

**Siamese网络训练**

代价函数：

The loss that is being minimized is then $L =$

$$\sum_{i}^{N}\left[\left\|f(x_i^a)-f(x_i^p)\right\|_2^2-\left\|f(x_i^a)-f(x_i^n)\right\|_2^2+\alpha\right]_+ . \qquad (1)$$

$\alpha$ is a margin that is enforced between positive and negative pairs. Superscript $a$, $p$, $n$ denote the anchor, positive and negative, respectively. The summation is over the set of all possible triplets in the training set.



A是某人，P是某人的另一张图，N是其他的人。

遍历所有三元组(A,P,N)，求其L的最小。公式中的参数α，是一个超参数，用于做margin，能避免模型输出的都是零向量。

用梯度下降法找到模型最优值。

## SRM滤波器

https://blog.csdn.net/c_chuxin/article/details/103981255

SRM指是《 Rich models for steganalysis of digital images》中提出来的，富隐写分析模型，34671维。论文中使用下面3个滤波器获得噪声图片：

$$\frac{1}{4}\begin{bmatrix}0&0&0&0&0\\0&-1&2&-1&0\\0&2&-4&2&0\\0&-1&2&-1&0\\0&0&0&0&0\end{bmatrix}\quad\frac{1}{12}\begin{bmatrix}-1&2&-2&2&-1\\2&-6&8&-6&2\\-2&8&-12&8&-2\\2&-6&8&-6&2\\-1&2&-2&2&-1\end{bmatrix}\quad\frac{1}{2}\begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\0&1&-2&1&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}$$

Figure 4. The three SRM filter kernels used to extract noise features.



## 论文整理

**假设**：自然图像不同子区域的噪声是相似的，但隐写图像对不同子区域进行修改，造成子**区域噪声不再相同**。

现有工作：

常见隐写方法：

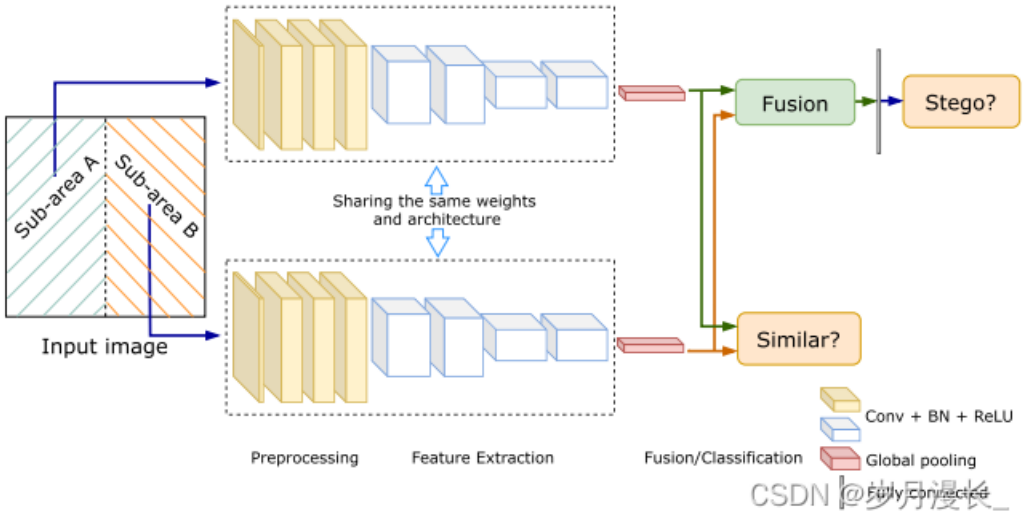
自适应隐写（最实用）：结合编码（STC）

基于深度学习的：综合型（GAN）；生成修改概率图；欺骗CNN隐写分析器；3-player对抗。

现有CNN隐写分析方法局限性：无法直接对大图像进行训练。

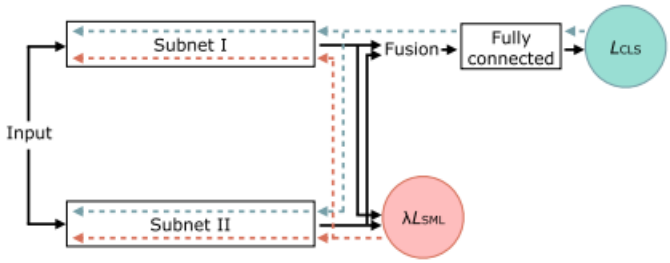
**目的**：针对任意大小图像进行隐写分析。

**提出方法**




把原始图像分成左右两部分，分别进行预处理、特征提取，得到两个信号，一个是分类/融合信号，用于判断是否为隐写图像，一个是相似度信号。
分类/融合信号：四个非线性矩阵：最大值，最小值，均值，方差。
方差和最大-最小的值在自然图像中更小，最大和最小值更接近均值

**代价函数：**

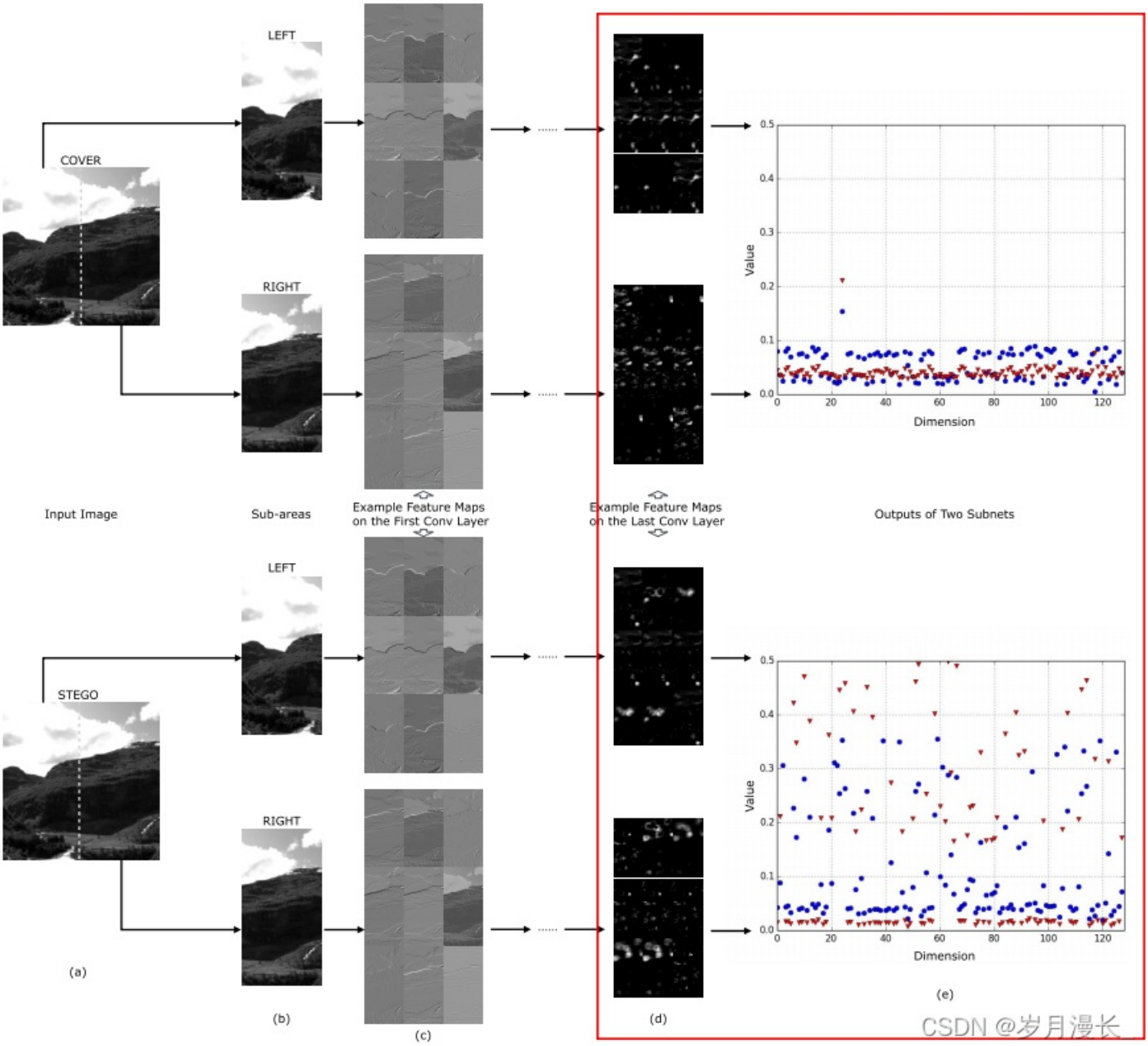



$$L = L_{\mathrm{CLS}}(p, y) + \lambda \cdot L_{\mathrm{SML}}(\mathbf{f}_{sub_i}, \mathbf{f}_{sub_j}, y)$$

论文中λ设为0.1，对比损失的m设为1，训练500epoch，15h

预处理：用5*5 的SRM kernels 进行卷积
特征提取：残差连接？最后获得一个128维的向量

实验结果：




## 代码运行bug：

1. `undefined symbol: PySlice_Unpack`（使用`conda install python==3.6`）版本不兼容
2. `no module named cv2`（使用`pip install opencv-python`）
3. `AttributeError: module 'torch._C' has no attribute '_cuda_setDevice'`
   （在python命令后面加上 --gpu_ids -1）
   参考https://blog.csdn.net/weixin_39450145/article/details/104797786
4. `os.makedirs(args.ckpt_dir, exist_ok=False)`
   报错：FileExistsError: [WinError 183] 当文件已存在时，无法创建该文件。: '...'
   把False改成True

能够成功运行，用MATLAB的WOW加密1000个图片（训练集：交叉验证：测试集=6:1:3），0.2bpp，训练迭代100次后正确率并不理想。（怀疑有其他问题，后续跟进）

```
2022-03-17 16:23:48.851 train.py:271[6002] INFO Epoch: 99
2022-03-17 16:23:48.851 train.py:272[6002] INFO Train
2022-03-17 16:24:17.187 train.py:274[6002] INFO Time: 2967.226026535034
2022-03-17 16:24:17.188 train.py:275[6002] INFO Test
2022-03-17 16:24:18.700 train.py:258[6002] INFO Test set: Loss: 0.7247, Accuracy: 54.02%)
2022-03-17 16:24:18.700 train.py:286[6002] INFO Best accuracy: 0.5625
2022-03-17 16:24:18.700 train.py:287[6002] INFO Time: 2968.739294528961
2022-03-17 16:24:18.801 train.py:271[6002] INFO Epoch: 100
2022-03-17 16:24:18.802 train.py:272[6002] INFO Train
2022-03-17 16:24:47.152 train.py:274[6002] INFO Time: 2997.1905875205994
2022-03-17 16:24:47.152 train.py:275[6002] INFO Test
2022-03-17 16:24:48.659 train.py:258[6002] INFO Test set: Loss: 0.7251, Accuracy: 54.46%)
2022-03-17 16:24:48.659 train.py:286[6002] INFO Best accuracy: 0.5625
2022-03-17 16:24:48.659 train.py:287[6002] INFO Time: 2998.698397159576
```