

基于半监督阶梯网络的红外乳腺癌检测方法研究

侯 丽 * 桂 伟

HOULi GUI Wei

摘 要

乳腺癌是除肺癌以外女性的第二大致死癌症,早期筛查对提高乳腺癌患者存活率至关重要。利用红外热图检测、定位和识别乳腺癌,具有直观、无损和能预测“未病”等优点,然而在临床应用中,人体红外热图不仅几何与灰度分辨率较低,而且,受成像环境和个体变化的影响较大,导致乳腺癌检测与分类的精度较低、假阳性率高。尤其医学类影像数据显得尤为匮乏,有标签的数据更是少之又少,特别是乳腺红外图片由于缺少公共数据库,因此数据问题成了提高检测精度的一大难题。本文尝试在小样本下用半监督阶梯网络方法对乳腺癌进行红外检测,对 94 个有标签数据增加 570 个无标签数据,并采用数据增强技术增加有标签数据体量,使用半监督阶梯网络在 batch Normalization 的帮助下,在测试集上得到了 89.55% 的分类精度。高于人类专家约 85% 的分类精度,对摆脱目前乳腺癌红外热图筛查中依赖专家读图的限制具有重要意义。

关键词

红外热像图;数据增强;半监督阶梯网络

doi: 10.3969/j.issn.1672-9528.2018.06.056

1 引入

红外热图在肿瘤还没有产生明确的形态学变化时,就可以通过乳房表面的温度变化检测出来,因此可比传统的黄金方法约提早十年检测到乳腺癌^[1]。此外红外法不需要接触人体,无射线入侵伤害,对 40 岁以下女性的紧致乳房、纤维囊性乳房、有埋植物的乳房检测亦有效,因此被广泛应用于乳腺疾病的筛查与疾病诊断。

如图 1 所示是一名患者的乳腺红外热图,左乳病变,右乳正常。明显可以看到两侧乳腺不对称的温度分布;左侧有以一团弥漫性的高温区;左侧乳腺下边缘有生理热区扩大,并呈白色,明显比右侧温度高。

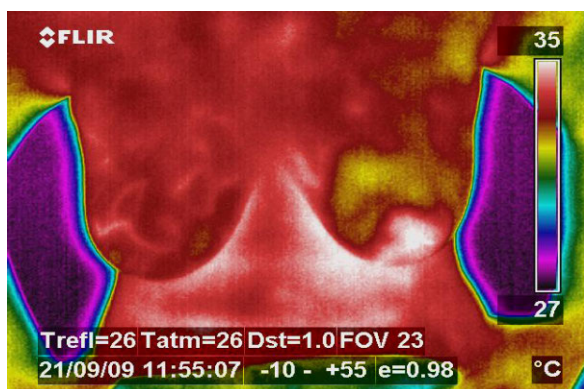


图 1 乳腺红外热图

在现有的文献中,对乳腺红外热图的分类,采用了不同的人工神经网络算法。Ng 和 Fok^[2]采用 MLP 网络,文献显示 61.5% 的准确性、40% 左右的敏感性和 68.9% 特异性。Ng 等人^[3]使用 RBF 网络的二进制图像分类,得到更好的结果:81.2% 的敏感性, 88.2% 的特异性和 80.9% 的精度。

Jakubowska 等人^[4]先分别应用 PCA 和 LDA 降维,降低对乳腺红外热图应用小波变换后计算的特征数量。特征维度降低后,用 MLP 神经网络对降维后的图片进行健康和病态的分类。

Koay 等人^[5]描述了一个用 BP 配置的人工神经网络(ANN),应用于五个特征子集的分类问题时不成功,后来减少为两个特征子集(平均温差和标准差),只使用一个隐藏节点后,结果显示只有两个假阴性。作者认为,假阴性的一个可能的解释为样本中癌症患者的图片较少(19 个样本中,有 14 个健康,只有 5 个异常样本)。为了充分评估 ANN 基于统计输入的分类能力,作者认为应当扩大样本量。文献^{[6][7]}也用 ANN 进行分类。

利用乳腺红外热图进行乳腺癌的检测面临着医学影像检测中常见的难题:数据不足问题,因此尚未有文献使用深度学习方法对其进行检测。本文所用图片取自巴西一个数据库^[8],节选其中专家诊断结果与图片可以对应的红外热图 47 幅,经过预处理,每幅热图分割为两个乳腺热图,因此得到 94 个样本,其中有 48 个健康样本,46 个异常样本。另外有 570

* 武汉商学院 湖北武汉 430056

个无标签样本。对于深度学习来说，样本量明显太小，小样本是否可以使用深度学习是目前争论非常激烈的一个话题，2017年5月31日 Jeff Leek 在 simplystatistics 上发表了一篇文章^[9]引起巨大反响，文章标题为“Don’t use deep learning your data isn’t that big”，本文借助于半监督阶梯网络，尝试在极小样本下用深度学习对乳腺癌进行红外检测实验。

实验在 Keras 上进行，Keras 是基于 TensorFlow 的一个深度学习框架（注：也可以支持 Theano，本文选用 TensorFlow 作为 Keras 的 backend），它的设计参考了 Torch，用 Python 语言编写，是一个高度模块化的神经网络库，支持 GPU 和 CPU。

2 半监督阶梯网络介绍

文献 [10] 将原有的阶梯网络从无监督学习应用到了半监督学习中。图 2 是阶梯网络的结构图，它在每一个编码层和对应的解码层之间都有一条横向的连接线，另外在编码层的每一层都加入高斯噪声，使得整个阶梯网络的结构图状如一层层的阶梯，阶梯网络也由此得名。对于无监督阶梯网络来说，在每一对编码层和解码层之间计算误差，使得每层对应的误差损失都尽量小，即可保证整个无监督网络丢失的信息尽量少。因此，将作为整个无监督阶梯网络的总损失。

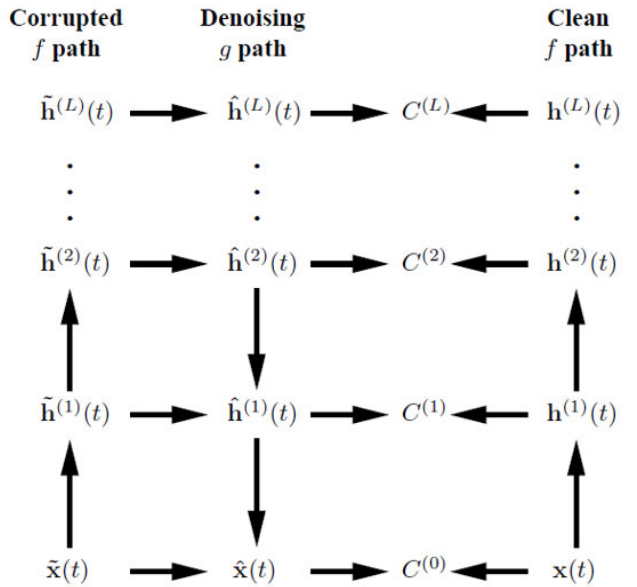


图 2 阶梯网络结构图

半监督阶梯网络是在无监督阶梯网络的基础上加上有监督学习，即在获得编码的最高层特征后接入有监督训练，并计算有监督损失函数。将无监督损失和有监督损失相加，作为整个半监督阶梯网络的总损失函数，利用梯度下降进行整体训练，达到既少丢弃信息，减小无监督损失，又借助高斯噪声，对每层编码高度抽象得到与分类相关的高级特征的目的。

的。

假设有 N 个有标签的数据 $(x(1), y(1))$ 、 $(x(2), y(2)) \dots$ ，及 M 个无标签的数据 $x(N+1)$ 、 $x(N+2)$ 、 $x(N+3) \dots M$ 远大于 N。希望通过这 M 个大量的无标签数据帮助少量的有标签数据学习，得到一个函数 $P(y|x)$ 用来对样本进行分类。在半监督阶梯网络中，对每个隐藏层加入噪声，然后结合有监督和无监督学习的损失函数作为整个阶梯网络的总损失函数，具体计算方法如公式 (1)

Cost = - \sum_{n=1}^N \log P(\tilde{y}(n) = y^*(n)|x(n)) + \sum_{n=N+1}^M \sum_{l=1}^L \lambda_l \text{ReconstructionCost}(\tilde{z}^{(l)}(n), \hat{z}^{(l)}(n))

X, y, y* 分别表示输入数据，无噪编码输出、添噪编码输出。每层的超参数 λ 表示每层编码解码损失函数的权重值，在建立模型时可以根据每一层的不同的重要性设置不同的权重值。具体结构图参见图 3。

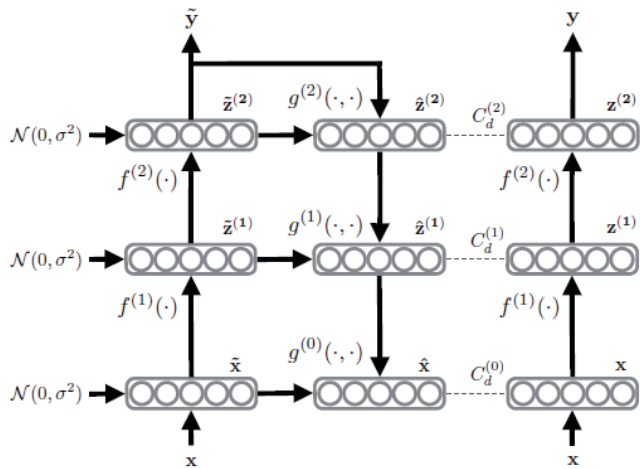


图 3 半监督阶梯网络

3 应用半监督阶梯网络的检测

将所有红外图片放在一起，共计 664 个样本，前 94 个为有标签数据，后面 570 个为无标签数据。

整个算法分为三部分：第一部分分别对输入数据进行清洁编码和添噪编码，两个编码过程基本相同，区别只在于添噪编码会在每一层添加上高斯随机噪声。对每一层的输入乘以权值（上标 l 表示第 l 层）以后运行 Batch Normalization^[11]，再对其通过变换重构，之后用 ReLU 函数激活作为下一层的输入。一层层迭代下来，在最后一层用 sigmoid 函数激活后，分别得到在原始数据下每一层的清洁编码和添噪编码。

第二部分是对无标签数据利用添噪编码的结果进行解码。从最后一层往前，每一层乘以权值后，进行批归一化，之后再高斯去噪作为下一层的输入，去噪后的结果通过该层的均值方差归一化得到每一层的解码结果。

第三部分是关于损失函数的度量。半监督阶梯网络的损失函数包含无监督损失和有监督损失两部分，无监督损失用每一层对应的添噪编码解码后的结果减去清洁编码的结果，再取平均得到，在每层间构建阶梯；有监督损失是对添噪编码的有监督样本段接入分类器，分类结果与实际标签计算交叉熵取平均值得到的。

具体算法流程参见伪代码（如图4）

Algorithm 1 Calculation of the output y and cost function C of the Ladder network

Require: $x(n)$

Corrupted encoder and classifier

$$\tilde{h}^{(0)} \leftarrow \tilde{z}^{(0)} \leftarrow x(n) + \text{noise}$$

for $l = 1$ **to** L **do**

$$\tilde{z}^{(l)} \leftarrow \text{batchnorm}(\mathbf{W}^{(l)} \tilde{h}^{(l-1)}) + \text{noise}$$

$$\tilde{h}^{(l)} \leftarrow \text{activation}(\gamma^{(l)} \odot (\tilde{z}^{(l)} + \beta^{(l)}))$$

end for

$$P(\tilde{y} | x) \leftarrow \tilde{h}^{(L)}$$

Clean encoder (for denoising targets)

$$h^{(0)} \leftarrow z^{(0)} \leftarrow x(n)$$

for $l = 1$ **to** L **do**

$$z_{\text{pre}}^{(l)} \leftarrow \mathbf{W}^{(l)} h^{(l-1)}$$

$$\mu^{(l)} \leftarrow \text{batchmean}(z_{\text{pre}}^{(l)})$$

$$\sigma^{(l)} \leftarrow \text{batchstd}(z_{\text{pre}}^{(l)})$$

$$z^{(l)} \leftarrow \text{batchnorm}(z_{\text{pre}}^{(l)})$$

$$h^{(l)} \leftarrow \text{activation}(\gamma^{(l)} \odot (z^{(l)} + \beta^{(l)}))$$

end for

Final classification:

$$P(y | x) \leftarrow h^{(L)}$$

Decoder and denoising

for $l = L$ **to** 0 **do**

if $l = L$ **then**

$$u^{(L)} \leftarrow \text{batchnorm}(\tilde{h}^{(L)})$$

else

$$u^{(l)} \leftarrow \text{batchnorm}(\mathbf{V}^{(l+1)} \hat{z}^{(l+1)})$$

end if

$$\forall i: \hat{z}_i^{(l)} \leftarrow g(\tilde{z}_i^{(l)}, u_i^{(l)})$$

$$\forall i: \hat{z}_{i,\text{BN}}^{(l)} \leftarrow \frac{\tilde{z}_i^{(l)} - \mu_i^{(l)}}{\sigma_i^{(l)}}$$

end for

Cost function C for training:

$$C \leftarrow 0$$

if $t(n)$ **then**

$$C \leftarrow -\log P(\tilde{y} = t(n) | x(n))$$

end if

$$C \leftarrow C + \sum_{l=0}^L \lambda_l \left\| z^{(l)} - \hat{z}_{\text{BN}}^{(l)} \right\|^2$$

图4 半监督阶梯网络的伪代码

将样本分为两部分，637个测试样本（其中570个无标签样本，67个有标签样本），27个测试样本。其中将测试样本进行了样本增强，提升至67个样本，提前保存在validation文件夹中，这样做一方面为了更好的检验模型的泛化性能，另一方面是为了将训练时和测试时的batch_size统一，方便编程。

模型各层编码过程及解码过程如图5所示：

```
In [3]: runfile('C:/Users/admin/.ipython/ladder_net_IR.py')
=== Corrupted Encoder ===
Layer 1 : 10000 -> 5000
Layer 2 : 5000 -> 1000
Layer 3 : 1000 -> 500
Layer 4 : 500 -> 250
Layer 5 : 250 -> 100
Layer 6 : 100 -> 2
=== Clean Encoder ===
Layer 1 : 10000 -> 5000
Layer 2 : 5000 -> 1000
Layer 3 : 1000 -> 500
Layer 4 : 500 -> 250
Layer 5 : 250 -> 100
Layer 6 : 100 -> 2
=== Decoder ===
Layer 6 : None -> 2, denoising cost: 0.1
Layer 5 : 2 -> 100, denoising cost: 0.1
Layer 4 : 100 -> 250, denoising cost: 0.1
Layer 3 : 250 -> 500, denoising cost: 0.1
Layer 2 : 500 -> 1000, denoising cost: 0.1
Layer 1 : 1000 -> 5000, denoising cost: 10.0
Layer 0 : 5000 -> 10000, denoising cost: 1000.0
```

图5 编码及解码示意图

运行结果显示，在阶梯网络框架下，没有用无标签数据

时分类精度约为49%~51%。

```
=== Training ===
Initial Accuracy: 50.7463 %
```

图6 监督分类结果

分别计算无监督损失和监督损失，共同作为模型的损失函数，用Adam优化器优化损失函数进行学习。分类结果如图7所示。

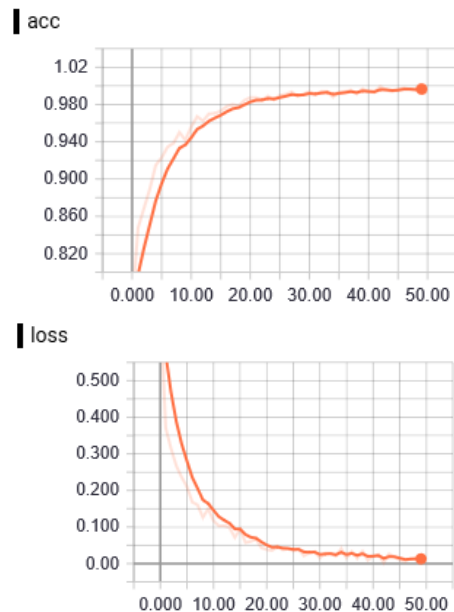




图7 半监督阶梯网络学习结果

从分类结果可以看出,模型收敛性很好,并且在训练集和测试集上分别达到了98.31%和89.55%的分类精度(见图8)。

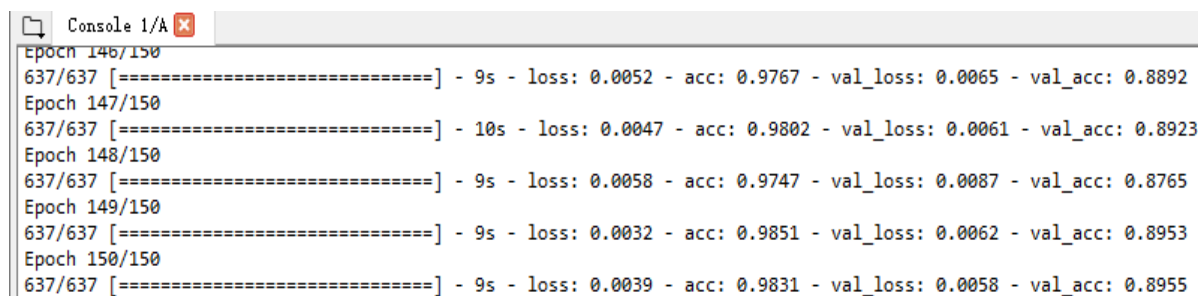


图8 半监督阶梯网络分类精度

4 总结

本文对有限的乳腺红外数据进行阶梯网络的半监督学习,并没有进行过多的调参,而且无标签样本数量也不多,已经得到89.55%的分类精度,说明深度学习有了一定的数据量以后,可以迅速提高学习表现,分类精度可以轻易达到甚至超越浅层学习的结果。

鉴于红外疾病检测中,都是疾病导致体表温度变化,从而引起红外热图的改变,不同疾病在红外热图上都有类似的表现,图片具有极强的关联性,因此基于半监督阶梯网络的红外乳腺癌检测方法可以对其他疾病的红外检测有很好的迁移性。

参考文献:

- [1] E.Y.K. Ng, A review of thermography as promising noninvasive detection modality for breast tumor, International Journal of Thermal Sciences 48 (1) (2009) 849 - 859.
- [2] E.Y.K. Ng, S.C. Fok, A framework for early discovery of breast tumor using thermography with artificial neural network, The Breast Journal 9 (4) (2003) 341 - 343.
- [3] E.Y.K Ng, E.C. Kee and R. Acharya, (2005) Advanced technique in breast thermography analysis, Proceedings of IEEE Conference on Engineering in Medicine and Biology, pp. 710 - 713.

[4] T. Jakubowska, B. Wiecek, M. Wysocki, C. Drews-Peszyński and M. Strzelecki, (2004) Classification of breast thermal images using artificial neural networks, Proceedings of 26th Annual IEEE EMBS Conference, 1155 - 1158.

[5] J. Koay, C. Herry, M. Frize, Analysis of breast thermography with an artificial neural network, Engineering in Medicine and Biology Society—IEMBS 1 (1) (2004) 1159 - 1162.

[6] Wakankar A T, Suresh G R. Automatic Diagnosis of Breast Cancer using Thermographic Color Analysis and SVM Classifier[J]. 2016.

[7] Lessa V, Marengoni M. Applying Artificial Neural Network for the Classification of Breast Cancer Using Infrared Thermographic Images[M]// Computer Vision and Graphics. Springer International Publishing, 2016.

[8] Visual Lab. <http://visual.ic.uff.br/dmi/>.

[9] Jeff Leek. Don't use deep learning your data isn't that big. <https://simplystatistics.org/2017/05/31/deeplearning-vs-leekasso/>

[10] Rasmus A, Valpola H, Honkala M, et al. Semi-Supervised Learning with Ladder Networks[J]. Computer Science, 2015, 9 Suppl 1(1):1-9.

(收稿日期: 2018-06-12)