

基于密集连接神经网络的肺部X光 多分类方法

陈 财*, 王世界, 朱习军

青岛科技大学, 信息科学技术学院, 山东 青岛

收稿日期: 2022年3月7日; 录用日期: 2022年5月9日; 发布日期: 2022年5月16日

摘 要

近几年, 神经网络在医学病种分类方面经历着迅速的发展, 在某些病种上面的分类水平已经达到甚至超过专业医学从业人员的水准。但是神经网络在医学上的应用主要集中在病种的二分类方面, 在病种的多分类方面神经网络模型的表现还不尽人意。概括的说, 神经网络特征提取的不完美、医学图像的复杂性以及相关领域专业知识的匮乏是导致该结果的三个主要因素。为此, 本文提出了一种基于神经网络密集连接并且包含置信度概念的多分类方法。该方法能够有效融合病种的多维度特征并且使用置信度概念保证分类结果的可靠性。本文在CXRI4数据集上进行测试, 实验表明, 该方法能够有效融合不同维度的特征, 从而增强了特征提取的能力; 引入置信度概念的分类方法应用于病种标签树能够获取达到预期水平的分类可信度, 而不是像目前的二分类只能提供绝对分类结果。这为临床医疗诊断提供了更加丰富和可靠的病种分类信息。

关键词

神经网络, 多层分类, 置信度, 密集连接

Lung X-Ray Multi-Label Classification Method Which Is Based on Densely Connected Neural Network

Cai Chen*, Shijie Wang, Xijun Zhu

College of Information Science and Technology, Qingdao University of Science and Technology,
Qingdao Shandong

Received: Mar. 7th, 2022; accepted: May 9th, 2022; published: May 16th, 2022

*第一作者。

Abstract

In recent years, neural networks have experienced rapid development in the medical diseases classification. The classification ability of certain diseases has reached or even exceeded the level of professional medical practitioners. However, the application of these neural networks is mainly focused on the binary classification of diseases, and the performance of neural network models in the multi-label classification of diseases is not satisfactory. The imperfection of feature extraction in the neural network, the complexity of medical images, and the lack of expertise in related fields are the three main factors leading to this result. To this end, this paper proposes a confidence multi-label classification method based on dense neural network connections, which can effectively fuse the multi-dimensional features of the disease and use the concept of confidence to ensure the reliability of the classification results. This article is tested on the CXR14 datasets. Experiments show that the method in this article can effectively integrate features of different dimensions to enhance the ability of feature extraction. Based on the concept of confidence, we can obtain the expected level of classification credibility instead of the current definite classification that can only provide absolute classification results. This provides more abundant and reliable identification information for clinical medical diagnosis.

Keywords

Neural Networks, Hierarchical Multi-Label Classification, Confidence Factor, Densely Connect

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当下随着社会福利的扩大以及医疗科技的发展,先进的医疗设备已经在医院中得到了普及,医疗病理图像的获取变得越来越简单和高效。例如 X 光片、CT 等甚至都可以实现即时成像。然而目前从医疗图像中获取病理数据信息还主要是依靠医学相关从业人员的手动标注[1]。因此,相对应的巨大工作量和提取低效率让人们不断寻找更合适的从图像中获取所需信息的手段。依赖于神经网络的发展,使用计算机自动提取医疗图像中的信息也正在经历着快速的进步并且在很多方面已经达到甚至超过了专业从业者的水平。例如深度学习在皮肤癌诊断[2]、组织轮廓提取[3]、肺部 X 光照片的肺炎检测[4]。本文就是将研究集中于基于神经网络的医疗图像病种分类。

然而使用神经网络进行病种分类中一个比较大的挑战是需要大量的标签信息来标注图像的病理位置和病症,相对于医疗领域的图像,这需要耗费大量的时间和精力[5] [6]。就此,目前神经网络医疗图像分类有两种解决办法,一是改进网络提高其特征呈现能力以使其降低对数据标注信息完备性的依赖;二是在不确定病症的前提下获得该病症的父类病症从而有利于接下来的排查工作和诊断治疗[7]。接下来的工作就是尝试从这两个角度出发,尝试设计一种融合多维特征并且加入置信度概念的多分类方法,来尽可能提供准确的病种多分类结果。

本文的主要贡献有两个部分:

- 1) 改进网络来提供更强的特征提取能力;
- 2) 提供置信度的概念来输出用户定义角度的分类结果,而不是让神经网络输出绝对的结果标签信

息。

2. 引言

问题定义

临床形成的医疗图像中会反映患者的所有病理特征，而这些病理特征不但是不局限于一种病种而且在一种病种时也会有分不同阶段病理特征的特点。目前神经网络应用于胸部 X 光片病理识别主要还是集中于单病种的二分类识别并且基于神经网络的分类方法还不能很好的融合多维度的特征，故对病症的多阶段特征提取能力还比较弱[8]。

比如现在做得比较好的胸部 X 光片肺结节检测，虽然已得到了较大的发展甚至在某些条件下神经网络的检测能力已经达到甚至超过了专业从业医师的水准[9]，但是在实际操作中医疗图像在输入网络前会进行预处理从而将检测范围限定到大于等于 3 mm 的肺结节水平上。因此在实际应用时还有较大的局限性。

目前对于神经网络模型的病种分类，基本流程是确定要使用 backbone 网络用于特征信息，进一步，在此基础上加上各种模块[10]。对于一般的网络模型可以用以下公式表示：

$$z = Wx + b, o = f(z)$$

其中 W 表示神经网络权重， b 代表偏置， x 代表输入特征， z 代表输出，函数 f 代表激活函数， o 表示该层神经网络的最终输出。

经过多层神经网络的特征提取后，将结果与对应数据的标签信息相匹配，并根据两者之间的差距计算损失，损失函数定义为：

$$Loss = J(f(xi), yi)$$

其中函数 f 用来计算数据在神经网络中的输出结果， $f(xi)$ 表示第 i 个对象的结果， J 表示计算神经网络输出结果与实际标签值 yi 之间的差值。

在计算完他们之间的差值后，就可以使用损失函数的梯度更新公式来更新网络权重。

在多分类网络中，目前最基本的办法是分别训练分类器，对所有的标签进行训练。这个过程相当于将二分类的方法应用于多标签分类中[7]。然而使用二分类的方法进行多分类会导致多方面的不利。比如计算资源的浪费、没有考虑标签之间的关联关系等[11]。

目前针对这种情况有两种解决办法，即在考虑多标签关联关系的基础上可以采取自顶向下和自底向上的分类方法[7]。鉴于自顶向下的分类方法所产生的分类结果是基于父类分类标签是正确的基础上的，而父类标签的分类结果在计算子类结点标签时是不会更新的，因此父类节点的分类错误将会导致依据该父类进行的分类完全错误[12]。因此本文采用自底向下的分类策略并且引入了置信度的概念。在分类时会首先计算病种标签树所有终端节点的标签概率并以此为依据计算其父类结点标签值的概率大小[13]。

3. 基于密集连接神经网络的肺部 X 光多分类

在本节中介绍所提出的融合多层特征和置信度概念的多标签分类方法的细节。

3.1. 密集连接神经网络

相关工作者针对一般的链式神经网络结构随着网络不断加深容易出现的梯度消失、梯度爆炸等问题，提出了正则化、BN 等方法来缓解随着网络加深产生的这些问题，但实际上并未根本解决问题[14]。而链式结构这一本质特点所带来的梯度不稳定情况也正是神经网络特征信息传递和提取不敏感的一个重要原因。

为解决这个问题, GoogLeNet 第一个引入了多分支网络结构的概念, 通过构建不同的 Inception 模块以扩大网络的宽度和深度[15]。Inception 模块中不同的分支能够同时进行多种操作, 从而对输入层的不同维度特征进行特征提取[16]。ResNet 网络[17]首次提出了跳跃链接思想, 浅层网络的特征可以通过跳跃链接直接传递到更深层次的网络层而不需要经历中间网络变化导致的梯度消失问题。

再进一步, 2017 年提出的 DenseNet 网络[18], 采用了密集连接的方式, 通过将网络中的每一层与模块中之前每一层建立连接的方式来增强网络接收和处理不同维度图像信息的能力, 然而这同时意味着在向网络输入数据时整个网络需要存储大量的中间特征。这会导致 DenseNet 网络需要花费大量的内存开销。针对这个问题, 相关工作者提出了共享存储空间的方式以较小的时间代价减少对内存的消耗[19]。

对于基本的 DenseNet 网络, 其基本架构如图 1 所示。图 1 介绍了 DenseNet 网络的基本结构。

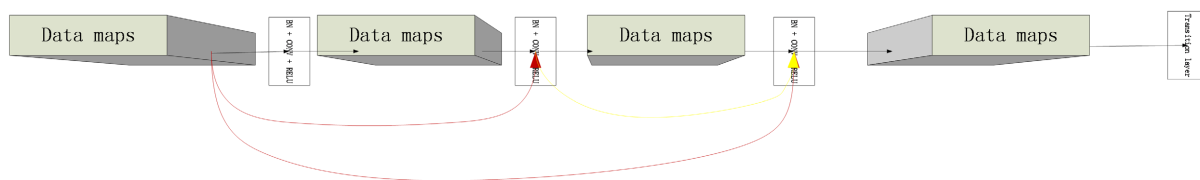


Figure 1. Example diagram of DenseNet network structure

图 1. 密集网络结构示例

为减少 DenseNet 网络的 GPU 显存消耗, 借鉴[19]采用了用较小的时间代价来换取空间消耗的方法。

具体来说, 在实验操作时, 针对 DenseNet 网络每一模块中的每一层都要获得之前所有层的特征的特点, 相对于在每次进行特征融合前为该次特征融合用到的所有层分配存储空间, 本文借鉴共享存储空间的原则来使用指针引用之前分配空间的之前层的网络特征图。相同的思想方法也会应用于接下来的归一化和卷积操作。具体操作如图 2 所示。

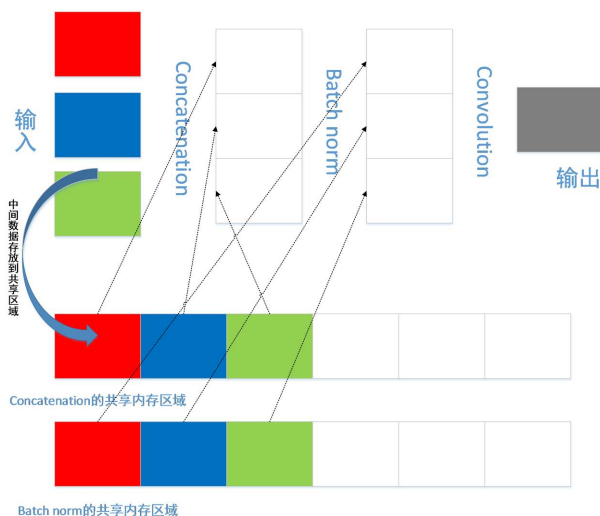


Figure 2. Principle of DenseNet network based on shared storage space

图 2. 密集网络共享参数示例图

对于梯度更新操作, 鉴于在共享存储空间中存放的是每一层神经网络的输出结果, 因此对于特征融合、归一化、卷积这些神经网络权重更新的操作未在共享存储空间中体现。在进行梯度计算时需要对这

些神经网络更新的权重重新计算,以获得其当下值后再计算梯度。这也是改进的 DenseNet 网络减少内存消耗所付出的主要时间代价。实验表明相比于不改进的 DenseNet 网络,所付出的时间代价不会超过 20%。

3.2. 基于置信度的多分类策略

在本节将详细介绍基于置信度的多分类策略。首先是需要构建分类的树状结构,其中根节点是异常,其下是各种病症,并且让父病症包括子病症。基于 CXR14 数据集构建的结构图如图 3:

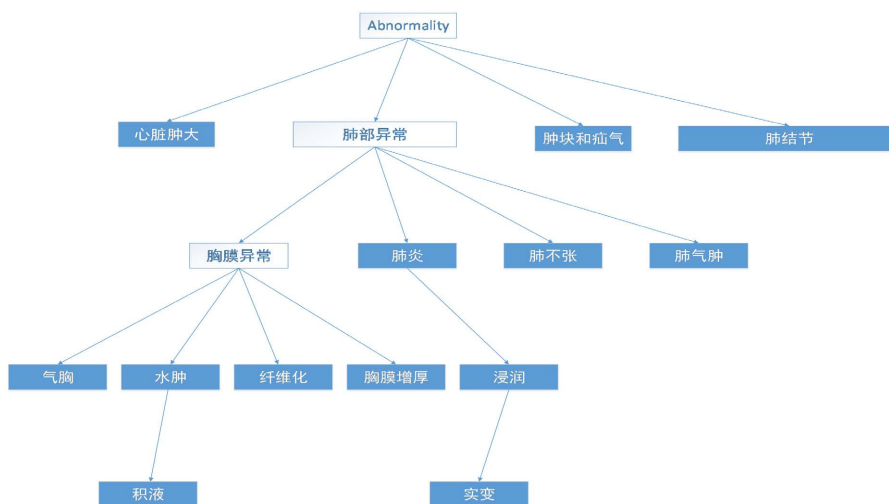


Figure 3. Concept Tree
图 3. 标签树结构

在训练过程中,首先需要统计所有节点的先验概率密度分布[20]。即统计所有的标签和每一个标签在所有标签中出现的概率。然后结合后验概率进行统计计算标签向量的概率分布。

具体来说,若某个实例 x 的实际标签是 l , 而使用分类器输出的结果标签向量是 s , 进一步将结果标签向量的值压缩到直方图 $H_l^+[s_q]$ 中。同样,还需要计算实例 x 不属于标签 l 的直方图值 $H_l^-[s_q]$ 。

对于 l 标签的父节点,假设用 a 表示,同样也需要计算器概率密度分布。为计算 a 的值,需要统计所有属于 a 的子节点的概率和并同样用向量的形式放到 s_q 中。进一步也要计算 a 的直方图 $H_a^+[s_q]$ 。对于 l 标签中的任意父节点标签,需要重复这个计算过程直至所有父节点计算完成。

当计算完所有节点后,对实例 x 关于 l 标签的先验概率可以用以下公式表示:

$$P(s_q | l) = H_l^+[s_q] / |H_l^+[s_q]|_l$$

$$P(s_q | \neg l) = H_l^-[s_q] / |H_l^-[s_q]|_l$$

最后计算 l 标签的后验概率 $P(l|s_q)$, 公式为:

$$P(l|s_q) = P(s_q | l)P(l) / (P(s_q | l)P(l) + P(s_q | \neg l)P(\neg l))$$

具体流程如算法 1 所示:

- ① 建立标签树
- ② 计算标签的先验概率分布
- ③ 对分类器输出的当前实例 x 进行一下操作
- ④ 获得实例 x 的标签值 l 并为实例标签中的每个标签赋值,压缩到 S_q 中,进一步转化为直方图形式

$$H_1^+[s_q]$$

- ⑤ 计算标签列表 1 中所有标签的父节点并汇总到 $H_a^+[s_q]$ 中
- ⑥ 计算所有不属于实例 x 标签列表 1 的标签值并汇总到 $H_1^-[s_q]$ 中
- ⑦ 通过之前计算的标签列表 1 的值计算器后验概率分布。

以上便完成了相对于实例 x 的相关标签 1 的概率分布的计算。接下来便是与事先设定的置信度阈值比较，来判断当前分类器对实例 x 的判断是否可信。

具体算法如下：

- ① 获得实例的后验概率分布 $P(l|s_q)$ ，并用符号 C 代表
- ② 判断 C 与阈值 T 的大小，若 C 小于阈值 T 则执行以下操作
- ③ 获得标签列表 1 的父节点并计算其按照算法 1 计算加入父节点的后验概率 C_+ ，再与阈值 T 相比较。

4. 实验结果与分析

4.1. 实验数据集

本文实验使用 CXR14 数据集作为测试数据集[21]。CXR14 数据集基于 NLP 从放射诊断报告书中获得标签数据，标签涵盖了 14 个病症，加上无异常，共 15 个标签。将肿块标签和疝气标签融合为一个，故分类时是有 13 个病症标签。数据包含 30,805 个病人的 112,120 张胸部前视 X 光图片。图 4 展示了病症分类以及互相关联关系。

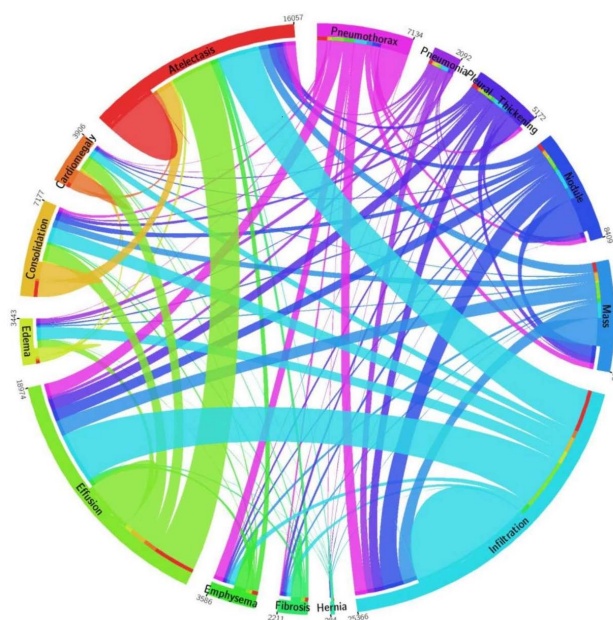


Figure 4. The proportions of images with multi-labels in each of 14 pathology classes and the labels' co-occurrence statistics

图 4. ChestXRay14 数据集中对各数据病种的介绍和病种之间的关联关系

本文选择数据中的 80% 作为训练数据集 D ，20% 作为测试数据集 T 。进一步在训练数据集中选取 20% 作为验证数据集 D 。

4.2. 评价指标

本文使用 AUC (Area under ROC curve)作为分类结果的评价指标，AUC 中涉及的基本概念如下：

TP (true positive): 分类正确，把原本属于正类的样本分成正类。

TN (true negative): 分类正确，把原本属于负类的样本分成负类。

FP (false positive): 分类错误，把原本属于负类的错分成了正类。

FN (false negative): 分类错误，把原本属于正类的错分成了负类。

ROC 的横坐标是为假阳性率(False Positive Rate, FPR)，是指在所有实际为阴性的样本中，被错误地判断为阳性之比率。其公式是：

$$FPR = FP / (FP + TN)$$

ROC 纵坐标是真阳性率(True Positive Rate, TPR)，是指在所有实际为阳性的样本中，被正确地判断为阳性之比率。其公式是：

$$TPR = TP / (TP + FN)$$

实际上 AUC 就是指 ROC 曲线下的面积大小，而一般认为 ROC 面积越大，分类器的分类性能就越强。

4.3. 实验结果分析

本文针对数据的多标签分类提出了一种基于置信度的多分类方法，骨干网络使用的是 densenet121 网络并使用了权值共享减少 GPU 显存消耗。Batchsizes 实验设置为 128，epoch 设置为 10。输入数据已预处理为 1024 * 1024 像素，在转入网络时将其处理为 224 * 224 像素的数据并随机进行翻转等微调操作。

本实验使用的平台是 pytorch1.7.1，GPU 是 NVIDIA RTX3090，损失函数使用了均方误差函数。梯度更新使用的是 Adam 算法并且学习率指定为 1e-3，冲量默认为 0.9。

图 5 显示了 AUC 的结果：

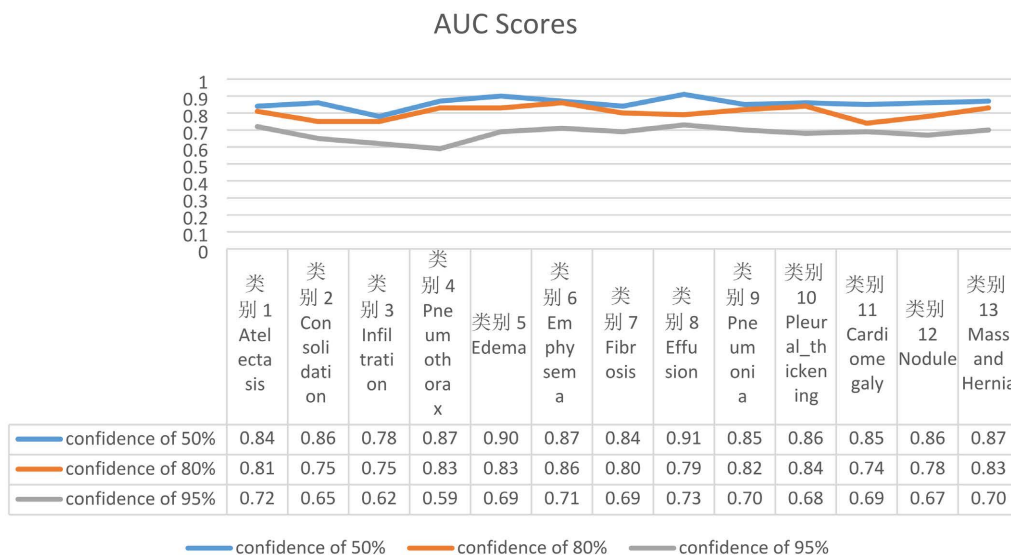


Figure 5. Comparison of AUC scores for all disease patterns with poch of 10

图 5. 病种分类的 AUC 成绩

如图 5 所示, 我们比较了三种置信度水平下的分类器的分类准确率。当置信度为 50% 时病种的平均分类准确率是 0.858; 置信度为 80% 时病种的平均分类准确率是 0.802; 当置信度为 95% 时病种的平均准确率为 0.68。

根据实验得到, 与不设置置信度概念的分类器相比, 在置信度为 50% 时分类器的分类准确率得到提高, 而在继续提高置信阈值时分类器的分类准确率有所下降。

5. 总结与展望

目前神经网络在医疗图像上的研究主要集中在二分类方面, 然而二分类在实际应用中存在着诸多问题。本文提出了一种基于密集连接网络且采用置信度概念的胸部 X 光多层多标签分类方法, 使分类的结果保证在一定置信度范围内, 并且父类标签的准确率普遍做到高于子类标签的。这种分类方法能够提供给使用者更有用和可靠的分类信息。因此具有很好的临床参考意义。然而目前分类中仍然存在很多不足, 例如置信度阈值需要用户给定, 而设置多大阈值最好仍不明确; 对标签进行划分时, 会有许多干扰因素, 如何合理的处理这些干扰因素仍需进一步研究。而这些也将是接下来的工作内容和方向。

参考文献

- [1] Varela, C., Timp, S. and Karssemeijer, N. (2006) Use of Border Information in the Classification of Mammographic Masses. *Physics in Medicine & Biology*, **51**, 425-441. <https://doi.org/10.1088/0031-9155/51/2/016>
- [2] Andre, E., Brett, K., Novoa, R.A., et al. (2017) Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature*, **542**, 115-118. <https://doi.org/10.1038/nature21056>
- [3] Zhu, W., Huang, Y., Zeng, L., et al. (2018) AnatomyNet: Deep Learning for Fast and Fully Automated Whole-Volume Segmentation of Head and Neck Anatomy. *Medical Physics*, **46**, 576-589. <https://doi.org/10.1002/mp.13300>
- [4] Rajpurkar, P., Irvin, J., Zhu, K., et al. (2017) CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.
- [5] Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press, Cambridge.
- [6] Song, L., Xu, Y., Zhang, L., et al. (2020) Learning from Synthetic Images via Active Pseudo-Labeling. *IEEE Transactions on Image Processing*, **29**, 6452-6465. <https://doi.org/10.1109/TIP.2020.2989100>
- [7] Chen, H., Miao, S., Xu, D., et al. (2020) Deep Hierarchical Multi-Label Classification Applied to Chest X-Ray Abnormality Taxonomies. *Medical Image Analysis*, **66**, Article ID: 101811. <https://doi.org/10.1016/j.media.2020.101811>
- [8] Zhang, M. and Zhou, Z. (2014) A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge & Data Engineering*, **26**, 1819-1837. <https://doi.org/10.1109/TKDE.2013.39>
- [9] Jaeger, S., Karargyris, A., Candemir, S., et al. (2013) Automatic Screening for Tuberculosis in Chest Radiographs: A Survey. *Quantitative Imaging in Medicine and Surgery*, **3**, 89.
- [10] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [11] De Mbczyński, K., Waegeman, W., Cheng, W., et al. (2012) On Label Dependence and Loss Minimization in Multi-Label Classification. *Machine Learning*, **88**, 5-45. <https://doi.org/10.1007/s10994-012-5285-8>
- [12] Wehrmann, J., Cerri, R. and Barros, R. (2018) Hierarchical Multi-Label Classification Networks. <https://doi.org/10.1145/3019612.3019664>
- [13] Liu, B., Sadeghi, F., Tappen, M., et al. (2013) Probabilistic Label Trees for Efficient Large Scale Image Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 23-28 June 2013, 843-850. <https://doi.org/10.1109/CVPR.2013.114>
- [14] Meng, Z.Y., et al. (2021) Deep Neural Architecture Search: A Survey. *Journal of Computer Research and Development*, **58**, 22-33.
- [15] Szegedy C, Liu W, Jia Y, et al. (2014) Going Deeper with Convolutions. <https://arxiv.org/abs/1409.4842v1>
- [16] Szegedy C, Liu W, Jia Y, et al. (2015) Going Deeper with Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [17] He, K., Zhang, X., Ren, S., et al. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778.

- <https://doi.org/10.1109/CVPR.2016.90>
- [18] Huang, G., Liu, Z., Laurens, V., *et al.* (2016) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 4700-4708.
<https://doi.org/10.1109/CVPR.2017.243>
- [19] Pleiss, G., Chen, D., Huang, G., *et al.* (2017) Memory-Efficient Implementation of DenseNets.
- [20] Davis, J., Liang, T., Enouen, J., *et al.* (2019) Hierarchical Semantic Labeling with Adaptive Confidence. *International Symposium on Visual Computing*, Lake Tahoe, 7-9 October 2019, 169-183.
https://doi.org/10.1007/978-3-030-33723-0_14
- [21] Wang, X., Peng, Y., Lu, L., *et al.* (2017) ChestX-ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2097-2106.
<https://doi.org/10.1109/CVPR.2017.369>