计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(4):117−122 [doi: 10.15888/j.cnki.csa.008446]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于注意力机制多任务的肺结节癌变风险判断①

王广涵, 程远志, 史　操, 许灿辉

(青岛科技大学 信息科学技术学院, 青岛 266061)
通信作者: 程远志, E-mail: yzcheng2007@163.com

摘　要: 对于 CT 影像中检测出的肺部结节, 需要自动判断其是否有癌变风险. 不同于大多数现有的研究方法只区分结节良恶性, 本文提出了一个基于注意力机制的多任务学习模型, 将与结节良恶性相关的语义特征属性一并判断输出, 通过判断 9 个结节特征 (对比度、分叶征、毛刺征、球形度、边缘、纹理、钙化程度、大小以及恶性程度) 的同时实现内在特征的共享, 以达到提高各子任务性能的目的. 选择视觉转换器 (ViT) 模型作为多任务共享特征提取层, 整体模型采用动态加权平均方法来对各子任务的 Loss 函数进行优化. 在 LUNA16 数据集上的实验表明, 该学习框架可以提升肺结节癌变风险判断的性能, 且同时对其他语义特征的判断也能提升结果的可解释性.

关键词: 肺结节; 癌变; 低剂量螺旋 CT; 多任务; 注意力机制; 计算机辅助诊断; 医学影像处理

引用格式: 王广涵,程远志,史操,许灿辉.基于注意力机制多任务的肺结节癌变风险判断.计算机系统应用,2022,31(4):117–122. http://www.c-s-a.org.cn/1003-3254/8446.html

Risk Assessment of Lung Nodule Canceration Based on Attention Mechanism and Multitask

WANG Guang-Han, CHENG Yuan-Zhi, SHI Cao, XU Can-Hui

(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: For pulmonary nodules detected in computed tomography (CT) images, it is necessary to automatically determine whether they are at the risk of canceration. This study proposes a multitask learning model based on the attention mechanism. Different from most existing research methods which only distinguish between the benignity and malignancy of nodules, the proposed model also assesses and outputs the semantic features related to the benignity and malignancy of nodules. The assessment of nine nodule features (subtlety, lobulation, spiculation, sphericity, margin, texture, calcification, diameter, and malignancy) and the sharing of internal characteristics are conducted at the same time to improve the performance of each subtask. The vision transformer (ViT) model is selected as the multitask shared feature extraction layer, and the whole model uses the dynamic weighted average method to optimize the Loss function of each subtask. Experiments on the LUNA16 dataset show that the proposed learning framework can improve the risk assessment of pulmonary nodule canceration and that the assessment of other semantic features can also enhance the interpretability of the results.

Key words: lung nodule; canceration; low dose spiral CT (LDCT); multitask; attention mechanism; computer aided diagnosis; medical image processing

肺癌是全球范围内发病率和死亡率增长最快的恶性肿瘤之一, 病死率在恶性肿瘤中居首位[1]. 高病死率的原因之一就是肺癌患者早期症状不明显, 然而等到发病明显时, 再检查就已经是中晚期, 难以有效的治疗.

① 基金项目: 国家自然科学基金 (61806107, 61973180, 62002190)
收稿时间: 2021-07-07; 修改时间: 2021-08-11; 采用时间: 2021-08-17; csa 在线出版时间: 2022-03-22

Orlacchio 等人[2] 指出, 如果可以早期诊断, 进行肺癌切除手术后, 患者 5 年生存率可达到 40%~80%. 因此若能在肺癌早期阶段进行正确诊断和治疗, 将显著改善患者预后, 降低肺癌病死率, 早期肺癌没有明确转移, 可以得到几乎根治的效果.

肺癌的确诊需要进行穿刺或者活检, 但是这两项检查需要进行手术操作, 对医生要求比较高, 且耗时耗费, 一般只有通过影像检查确定癌症风险和恶性结节位置之后才会视情况进行病理诊断确认. 计算机断层扫描技术 (computed tomography, CT) 是目前临床上肺癌诊断最常用的影像检查手段.

想要尽早的发现癌症风险, 则需要经常性的进行检查, 低剂量螺旋 CT 扫描 (LDCT) 在显著减低辐射剂量的同时可保证较好的图像质量, 使得肺癌筛查也可以成为日常体检的项目之一, 也就能及时进行早期肺癌诊断[3].

其实际应用价值, 使得国内外的 9 部肺癌筛查指南/共识都推荐高危人群定时采用 LDCT 进行筛查[4], 但是目前 LDCT 肺癌筛查还未大规模普及应用是因为尚存在两个问题: 放射诊断医师的培养数量远跟不上需要分析的 CT 影像的增长量, 且人工分析的工作量也十分巨大, 时间和经济上的成本居高不下; 放射诊断医师人工观察 CT 图像给出的结果也还存在假阴性和假阳性而导致的漏诊、延误. 因此一种能够根据 CT 图像自动得出肺结节癌变良恶性风险的计算机辅助诊断系统 (CAD) 就显得十分必要.

近年来的研究中, 根据算法可以分为两类: 基于影像语义特征的方法和基于深度学习的方法.

前一种方法是通过机器学习算法来提取肺结节明确的影像组学特征, 如纹理特征、灰度特征、形态学特征等, 然后根据这些语义特征来得出肺结节的良恶性的分类结果. 比较有代表性的有: Suzuki 等人[5] 提出的训练多个神经网络, 采用 3 个灰度特征、2 个边缘特征、1 个形状特征和临床信息用于良恶性肺结节的分类方法; 张佳嘉等人[6] 也是用同一数据集训练出多个分支网络, 先提取出医生诊断时所用的肺结节语义特征信息, 然后用于完成肺结节良恶性诊断任务. 这种方法的优点在于能够给出病理诊断的依据, 可解释性上更强一些.

后一种方法是直接应用深度学习等方法, 不去关心放射科医师的诊断依据, 充分利用 CT 图像的全部特征信息, 直接得出良恶性判断的结果. 比较有代表性的有: Ardila 等人[7] 提出了一种端到端的肺结节检测和肺癌风险预测模型, 利用 3D CNN 网络, 可以对同一病患的多张不同时期的 CT 影像进行对比处理, 检测出肺结节区域后结合 CT 全局信息给出一个恶性肿瘤得分, 取得了很好的效果; Hua 等人[8] 提出了两种深度学习架构, 来对肺结节的良恶性直接进行分类, 规避了对语义特征进行细化的需要, 也取得了不错的分类性能.

在对比过两种方法的优缺点之后, 本文提出了一种基于多头注意力机制的多任务学习方法, 既采用深度学习网络来充分提取 CT 图像特征, 又利用硬参数共享的多任务模式来得出结节良恶性以及相关语义特征, 避免了两阶段判断网络受到第一步分析识别各语义特征任务性能影响的情况, 同时还能利用关联语义特征提升对于良恶性判断的性能.

方法流程如图 1 所示, 从原 CT 图像中根据检测出的结节坐标, 将需要判断的结节截取出来, 然后进行图像处理输入到基于多头注意力机制的多任务学习网络中, 进而得出良恶性程度等语义属性的判断结果.

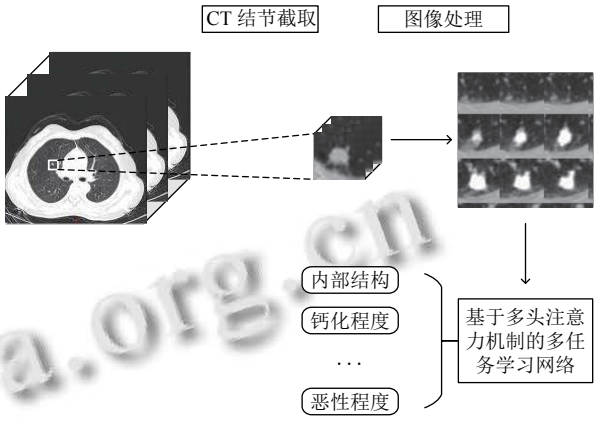

图 1 整体流程图

1 多任务框架

多任务学习 (multitask learning, MTL) 的概念最早由 Caruana[9] 在 1997 年提出, 他构造了一个前馈神经网络, 提出了一种通过训练单一的多层的感知器来执行多个任务的机制, 在辅助医疗诊断和自动驾驶等领域得到了成功应用.

选择多任务学习是因为单任务只关注一个目标而可能忽略了更多有用的信息. 具体来说, 这些信息来自于相关任务的训练信号. 通过在相关任务之间共享表示, 我们可以使我们的模型能够更好地对原始任务进行泛化[10]. 多任务学习的目的是在一并解决多个相关

任务的同时实现内在特征的共享. 已经证明, 这种共享可以提高部分或所有任务的性能[11].

目前基于深度神经网络的多任务学习中常用的模式有两种: 硬参数共享与软参数共享, 相对软参数共享模式来讲, 硬参数共享模式不需要对每个任务都构建和训练单独的模型, 且可以降低过拟合的风险[12], 因此本文选择硬参数共享的多任务模式.

1.1 肺结节各属性特征

LIDC-IDRI[13] 是由美国国家癌症研究所发起收集的肺结节公开数据集, 用于高危人群早期肺癌的检测诊断. 除了影像数据还包含了 4 位胸部放射科医师的诊断结果 XML 文件. 医师针对每个结节给出 9 个医学语义特征 (对比度、分叶征、毛刺征、球形度、边缘、纹理、内部结构、钙化程度和恶性程度) 的具体分级 (1–5 级, 钙化程度为 1–6 级), 等级越高, 语义特征越明显.

LUNA16[14] 是该数据集的子集, 删除了 LIDC-IDRI 中切片厚度大于 2.5 mm 和肺结节小于 3 mm 的 CT 影像. 相对于原数据集, LUNA16 数据集能够更好的对计算机辅助诊断系统的性能进行评估. 其中包含 888 张肺部 LDCT 扫描数据, 以及 1 186 个肺结节的坐标位置和直径大小.

结合 LIDC-IDRI 数据集中 4 位胸部放射科医师的诊断结果, 就可以得到 LUNA16 数据集中全部肺结节的语义特征, 我们取 4 位医师的诊断结果的算术平均作为最终各语义特征的分级数值.

1.2 各属性特征与癌变风险相关关系

在构造多任务模型之前, 我们要先探讨其他语义特征与结节良恶性之间的相关关系, 以此来估计添加到多任务学习中的特征是否会对良恶性判断产生正向影响, 对此我们采用余弦相似度来衡量每一个语义特征 Y 与结节良恶性等级 X 的关系, 计算公式如下:

$$c(X,Y)=\frac{X\cdot Y}{|X||Y|}=\frac{\sum_{i=1}^{n}X_iY_i}{\sqrt{\sum_{i=1}^{n}X_i^2}\sqrt{\sum_{i=1}^{n}Y_i^2}} \tag{1}$$

计算结果如图 2 所示, 通过其他 8 个语义特征和结节大小与良恶性的相似系数, 可以发现: 除内部结构这一特征外, 其他语义特征与结节良恶性之间都存在较强的相关性, 因此我们选择这些语义特征作为多任务学习中的辅助任务.

1.3 多任务损失函数权重优化方法--动态加权平均

对于多任务的 loss, 最简单的方式是直接将这两个任务的 loss 直接相加, 得到整体的 loss. 这种 loss 计算方式的不合理之处是显而易见的, 不同任务 loss 的量级很有可能不一样, loss 直接相加的方式有可能会导致多任务的学习被某个任务所主导或学偏. 当模型倾向于去拟合某个任务时, 其他任务的效果往往可能受到负面影响, 效果会相对变差. 因此需要对每个任务的 loss 进行加权[15].

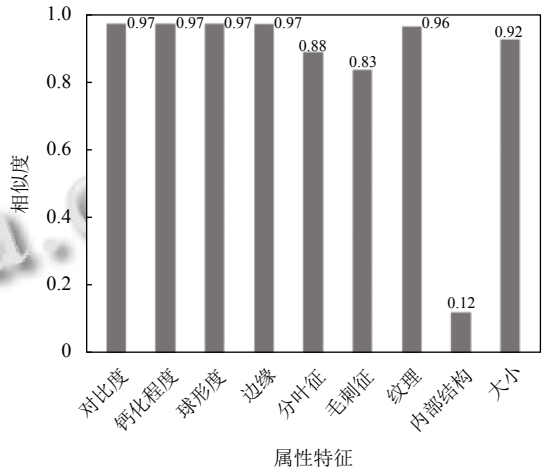

图 2 与良恶性相关关系

通过上一步得出的 8 个语义特征以及良恶性, 每个结节有 9 个标签来对应 9 个特征的平均等级评分, 每个子任务都有其损失函数, 如式 (2) 所示, 将这些子任务的损失函数加权求和便是整个网络的损失函数:

$$L_{\text{total}}(X,Y_{1:9})=\sum_{i=1}^{9}\lambda_iL_i(X,Y_i) \tag{2}$$

其中, $L_i(X,Y_i)$是以 X 为输入, Y_i为输出的第 i 个子任务的损失函数, 具体到语义特征属性的等级预测任务, 可以表示为式 (3):

$$\frac{1}{2n}\sum_{i=1}^{n}\left\|y_i-y_i'\right\|_2^2 \tag{3}$$

其中, n 为结节数量, y_i是第 i 个结节的真实值, y_i'是网络输出的预测值.

多任务学习中, 不同子任务的收敛速度, 训练难度都是不同的, 不能让简单任务主导整个训练, 导致各个子任务的表现差距过大, 对于大多数多任务网络而言, 训练过程中的最大难题是为每一个子任务找到合适的权重, 让每个子任务的重要性得到平衡, 解决这些问题的办法有几类, 有代表性的有: 梯度归一化 (GradNorm)[16], 使不同的任务 loss 量级接近; 动态任务优先级 (dynamic task prioritization)[17], 利用不确定性赋权值, 根据任务难易程度进行赋权值等.

本文采用的权重设计机制为动态加权平均 (dynamic weight average)[18], 通过考虑每个任务损失的变化率来调整任务权重. 受梯度归一化方法的启发, 动态加权平均方法对每个子任务首先计算前一个 epoch 对应损失的比值, 然后除以一个固定的值 T 进行 exp 映射后, 计算各个损失所占比. 如式 (4) 所示, 首先计算一个 epoch 后的损失变化:

$$\omega_i(t-1)=\frac{L_i(t-1)}{L_i(t-2)} \tag{4}$$

然后, 将$\omega_i(t-1)$带入式 (5), 得到对应子任务 i 的权重:

$$\lambda_i(t)=\frac{9\exp(\omega_i(t-1)/T)}{\Sigma_i\exp(\omega_i(t-1)/T)} \tag{5}$$

其中, T 的大小代表了任务间的松散程度, 如果该值足够大, 那么λ便会趋向于 1, 代表各任务的权重相同. 得到各任务权重之后, 就按照整体损失函数对共享网络结构进行优化调整.

2 基于多头注意力机制的多任务学习框架

Transformer 是 Vaswani 等人[19] 于 2017 年提出的一种模型架构. 它开创性的思想, 颠覆了以往序列建模和 RNN 划等号的思路, 之后被广泛应用于自然语言处理的各个领域, BERT、GPT 等模型也都是基于 Transformer 的模型, 在自然语言处理各个任务中都取得了突破性的成果.

因其在自然语言处理领域的巨大成功, 开始有研究者尝试将 Transformer 引入到计算机视觉领域, 其中由 Dosovitskiy 等人[20] 提出的 Vision Transformer (ViT) 模型很好的保留了 Transformer 的原始框架, 在图像分类任务上可以获得与当前最优卷积网络相媲美的结果.

并且因为 Transformer 中所应用的多头注意力机制, 将模型分为多个头, 形成多个子空间, 可以让模型去关注不同方面的特征/信息, 正适用于多任务学习中对不同的子任务关注不同的特性, 因此本文采用 ViT 模型作为多任务学习框架的共享特征提取层.

2.1 多任务学习分类框架

本文的多任务学习方法框架, 如图 3 所示.

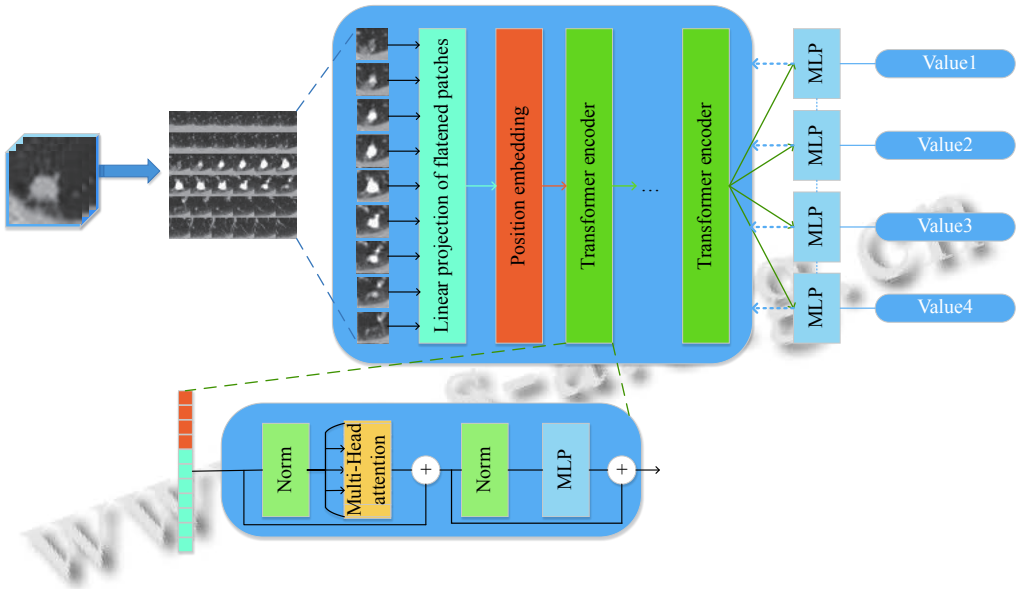

图 3 模型框架图

(1) 根据检测出的结节坐标, 从 CT 原图上取出小块, 然后采用了一种保留 3 维信息的 CT 图像转换 ViT 切片数据的方法, 将 3 维数据通过切片拼接转为符合特征提取网络输入所需的二维图片形式, 来完成第一部分的图像处理任务.

(2) 将训练集数据输入基于 ViT 模型的多头注意力机制特征提取网络中, 提取出的特征向量作为后续不同的分类任务网络的输入.

ViT 模型为了将一个标准 Transformer 直接应用到图像上, 尽可能少的修改, 选择将图像分割成小块, 并将这些块转化为线性嵌入序列, 然后在图像 patch 的嵌入中加入位置嵌入, 通过不同的策略在全局范围内保留空间/位置信息. 之后因为是分类任务, 只需要输入 Transformer 的 Encoder 层就可以了, 而无需像自然语言处理的翻译任务中还需要加 Decoder 层.

本文中, 通过第一步图像处理的方法, 将 3 维 CT

图像的每一个截面切片作为一个 patch, 进行线性序列化后, 嵌入位置序列信息, 作为 Transformer 的输入, 更能够保留和获取空间/位置信息.

(3) 采用多个分类任务共享特征提取网络的硬参数共享模式, 将 ViT 模型提取出的特征向量分别输入不同的 MLP 网络, 对结节的多个属性进行良恶性的判断. 使用动态加权平均 (DWA) 的损失函数权重计算优化方法, 保证多个分类任务的 loss 同步降低, 反向调整特征提取网络以获取跟任务更为相关的特征, 动态整体优化各分类网络的准确性. 模型在共享层就要学到一个通用的嵌入式表达使得每个任务都表现较好, 从而降低过拟合的风险, 以达到通过多任务学习, 提升该模型的泛化效果.

2.2 保留 3 维信息的 CT 图像转换 ViT 切片数据的方法

本文选用基于 ViT 模型的多头注意力机制特征提取网络, 因为 Transformer 模型的特性, 加载在更大样本上预训练好的模型参数, 然后再根据具体的任务迁移到目标数据集进行调整, 可以获得更好的结果, 并且所需的计算资源大大减少. 然而 ViT 模型现有的预训练参数都是用二维图像数据训练得出的, 因此想要应用于 3 维的 CT 数据上, 需要对 CT 数据进行变形调整.

受到 ViT 对二维图像进行切片取 patch 方法的启发, 本文采用的方法是, 对 3 维的 CT 数据进行截面切片, 形成的切片大小可以对应 ViT 模型的 patch 大小, 然后将一整套切片图片拼接成一副二维图像, 就可以应用现有的 ViT 预训练模型进行迁移学习. 这样即最大程度上保留了 CT 数据的 3 维信息, 还可以保证 ViT 模型进行裁剪图像时, 取到的每一片 patch 都是原 CT 图像的一个完整的截面图像.

图 4 为拼接后的二维图像以及经过 ViT 的多头注意力机制处理后的注意力热力图, 可以看到注意力集中于结节区域.

3 实验分析

本文对 LUNA16 数据集上进行五折交叉验证实验, 结节各特征属性值按照第 1 节所述的, 取 4 个医师给定值的算数平均, 用两种方式分类: (1) 评级小于 3 的归为负样本, 评级大于 3 的归为负样本, 评级等于 3 (4 名医师全部认为该结节评级不确定) 的归为不确定样本, 进行三分类; (2) 评级小于 2.75 的归为负样本, 评级大于 3.25 的归为正样本, 排除掉最不确定的情况进行二分类.

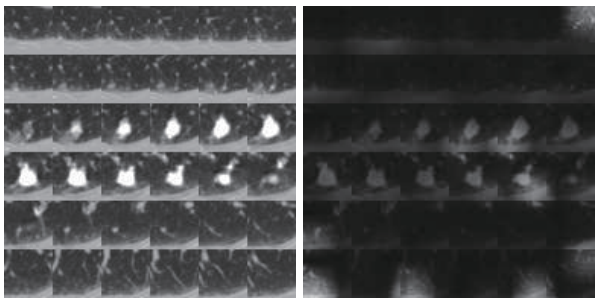

图 4 注意力热力图

分别以上述两种分类方式进行单任务和多任务的实验, 三分类的结节良恶性判断准确率远小于二分类的准确率, 此结果表明人类医师无法准确判断识别的结节, 使用本文提出的判断语义特征的多任务学习框架同样无法识别判断.

下面仅就二分类的实验结果进行具体说明.

3.1 单任务网络实验

首先用最基础的 MVC 模型, 通过第 1 节选定的 9 个语义特征的值来判断结节良恶性类别, 准确率达到了 93%, 符合我们对于语义特征与良恶性具有相关关系的分析.

我们使用 ViT 模型分别对每一个语义特征单独训练, 部分结果如表 1 所示.

表 1 单任务网络分类结果

语义特征	准确率	灵敏度	特异性
对比度	0.713	0.683	0.732
分叶征	0.841	0.832	0.847
毛刺征	0.837	0.853	0.827
钙化程度	0.839	0.819	0.852
恶性程度	0.814	0.803	0.821

可以发现单任务对各语义特征判断的效果并不是很好, 如果用各个单任务网络得出的属性值再来对良恶性进行分类, 则相当于对原输入数据增加 20%–30% 的噪声干扰, 最终的准确率甚至会低于单任务直接对良恶性判断的结果.

3.2 多任务网络实验

如表 2 所示, 在使用多任务网络对全部语义特征一起训练时, 因其内在特征的共享, 使得各任务或多或少的有了性能上的提升.

虽然对于结节良恶性分类的准确率最高只有 87.3%, 未达到 SOTA, 但是本文提出的框架除良恶性判断之外, 还可以一步同时输出结节的 9 个语义特征属性值, 提供了良恶性判断的依据与标准, 更具有实际应用价值.

表 2 多任务网络分类结果

语义特征	准确率	灵敏度	特异性
对比度	0.747	0.765	0.736
分叶征	0.877	0.883	0.874
毛刺征	0.883	0.882	0.890
钙化程度	0.859	0.854	0.863
恶性程度	0.873	0.872	0.875

4 结论与展望

在本文中, 我们提出了一种基于注意力机制的多任务学习框架, 该方法可以同时对 CT 影像中肺结节的良恶性等 9 类语义特征进行判断. 我们在 LUNA16 数据集上的 1 186 个 CT 结节进行了试验评估, 实验结果表明该方法对肺结节癌变风险的似然性预测是有效的. 虽然准确率没有达到目前 3D 预测模型的精度, 但是本文的方法同时给出了其他语义特征的预测值, 能够对良恶性的判断, 下一步需要进行的研究工作是: (1) 交叉验证每一项语义特征对良恶性结果的影响; (2) 进一步提升包括良恶性在内的各特征属性的预测准确率.

参考文献

1 王丽萍. 肺癌免疫治疗现状与展望. 中华实用诊断与治疗杂志, 2017, 31(2): 105–110.

2 Orlacchio A, Schillaci O, Antonelli L, *et al*. Solitary pulmonary nodules: Morphological and metabolic characterisation by FDG-PET-MDCT. La Radiologia Medica, 2007, 112(2): 157–173. [doi: 10.1007/s11547-007-0132-x]

3 颜小艳, 徐亮, 彭世秀, 等. 低剂量螺旋 CT 扫描在早期肺癌诊断中的应用价值. 海南医学, 2018, 29(7): 969–971. [doi: 10.3969/j.issn.1003-6350.2018.07.026]

4 石英杰, 李江, 孟耀涵, 等. 全球肺癌筛查指南及共识质量评价. 中华流行病学杂志, 2021, 42(2): 241–247. [doi: 10.3760/cma.j.cn112338-20200806-01035]

5 Suzuki K, Li F, Sone S, *et al*. Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. IEEE Transactions on Medical Imaging, 2005, 24(9): 1138–1150. [doi: 10.1109/TMI.2005.852048]

6 张佳嘉, 张小洪. 多分支卷积神经网络肺结节分类方法及其可解释性. 计算机科学, 2020, 47(9): 129–134. [doi: 10.11896/jsjkx.190700203]

7 Ardila D, Kiraly AP, Bharadwaj S, *et al*. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine, 2019, 25(6): 954–961. [doi: 10.1038/s41591-019-0447-x]

8 Hua KL, Hsu CH, Hidayati SC, *et al*. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. OncoTargets and Therapy, 2015, 8: 2015–2022.

9 Caruana R. Multitask learning. Machine Learning, 1997, 28(1): 41–75. [doi: 10.1023/A:1007379606734]

10 Ruder S. An overview of multi-task learning in deep neural networks. arXiv: 1706.05098, 2017.

11 Evgeniou T, Micchelli CA, Pontil M. Learning multiple tasks with kernel methods. The Journal of Machine Learning Research, 2006, 6: 615–637.

12 Baxter J. A Bayesian/information theoretic model of learning to learn via multiple task sampling. Machine Learning, 1997, 28(1): 7–39. [doi: 10.1023/A:1007327622663]

13 Armato III SG, Roberts RY, McNitt-Gray MF, *et al*. The Lung Image Database Consortium (LIDC): Ensuring the integrity of expert-defined "Truth". Academic Radiology, 2007, 14(12): 1455–1463. [doi: 10.1016/j.acra.2007.08.006]

14 Setio AAA, Traverso A, de Bel T, *et al*. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Medical Image Analysis, 2017, 42: 1–13. [doi: 10.1016/j.media.2017.06.015]

15 Vandenhende S, Georgoulis S, Van Gansbeke W, *et al*. Multi-task learning for dense prediction tasks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. [doi: 10.1109/TPAMI.2021.3054719]

16 Chen Z, Badrinarayanan V, Lee CY, *et al*. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018. 794–803.

17 Guo M, Haque A, Huang DA, *et al*. Dynamic task prioritization for multitask learning. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 282–299.

18 Liu SK, Johns E, Davison AJ. End-to-end multi-task learning with attention. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2018. 1871–1880.

19 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 6000–6010.

20 Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. Virtual Event, 2021.