

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2025 年 1 月 9 日 (09.01.2025)



(10) 国际公布号
WO 2025/007941 A1

(51) 国际专利分类号:
G06V 10/764 (2022.01)

(21) 国际申请号: PCT/CN2024/103726

(22) 国际申请日: 2024 年 7 月 4 日 (04.07.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310822213.1 2023年7月4日 (04.07.2023) CN

(71) 申请人: 阿里巴巴 (中国) 有限公司 (ALIBABA (CHINA) CO., LTD.) [CN/CN]; 中国浙江省杭州市滨江区长河街道网商路 699 号 4 号楼 5 楼 508 室, Zhejiang 310052 (CN)。

(72) 发明人: 夏英达 (XIA, Yingda); 美国马里兰州北贝塞斯达麦格拉斯大道 5410 阿罗伍德公寓 413, Maryland 20852 (US)。袁铭泽 (YUAN, Mingze); 中国北京市朝阳区望京东园四区 6 号楼金辉大厦, Beijing 100102 (CN)。张灵 (ZHANG, Ling); 美国马里兰州洛克威尔国会巷 210 号 201 室, Maryland 20852 (US)。姚佳文 (YAO, Jiawen); 中国北京市朝阳区望京东园四区 6 号楼金辉大厦, Beijing 100102 (CN)。裘茗烟 (QIU, Mingyan); 中国浙江省杭州市余杭区文一西路阿里巴巴西溪园区 969 号, Zhejiang 311121 (CN)。董和鑫 (DONG, Hexin); 中国北京市朝阳区望京东园四区 6 号楼金辉大厦, Beijing 100102 (CN)。周靖人 (ZHOU, Jingren); 中国浙江省杭州市西湖区三墩镇灯彩街 1008 号云谷园区, Zhejiang 310030 (CN)。吕乐 (LU, Le); 美国马里兰州贝塞斯达, Maryland (US)。

(54) Title: IMAGE CLASSIFICATION METHOD AND APPARATUS, AND METHOD AND APPARATUS FOR TRAINING IMAGE CLASSIFICATION MODEL

(54) 发明名称: 图像分类方法、训练图像分类模型的方法及装置

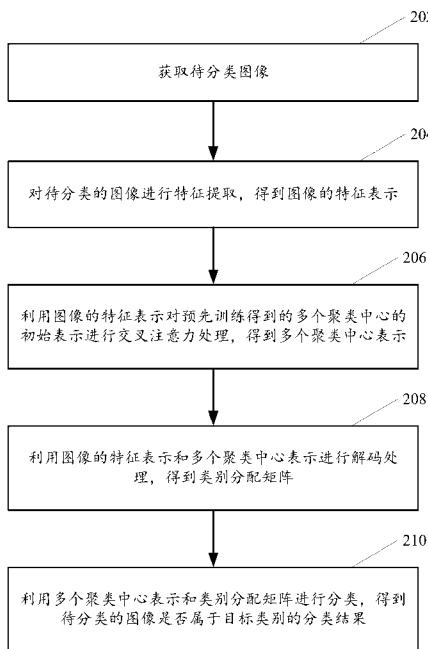


图 2

202 Acquire an image to be classified
204 Perform feature extraction on said image, so as to obtain a feature representation of said image
206 Use the feature representation of said image to perform cross attention processing on initial representations of a plurality of clustering centers which are obtained by means of pre-training, so as to obtain a plurality of clustering center representations
208 Perform decoding processing by using the feature representation of said image and the plurality of clustering center representations, so as to obtain a category allocation matrix
210 Perform classification by using the plurality of clustering center representations and the category allocation matrix, so as to obtain a classification result indicating whether said image belongs to a target category

(57) Abstract: Disclosed in the embodiments of the present application are an image classification method and apparatus, and a method and apparatus for training an image classification model. The main technical solution comprises: acquiring an image to be classified; performing feature extraction on said image, so as to obtain a feature representation of said image; using the feature representation of said image to perform cross attention processing on initial representations of a plurality of clustering centers which are obtained by means of pre-training, so as to obtain a plurality of clustering center representations; performing decoding processing by using the

(74) 代理人:北京同钧律师事务所(BEIJING TONGJUN LAW FIRM); 中国北京市海淀区西直门北大街32号院1号楼7层808室, Beijing 100082 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

— 关于申请人有权要求在先申请的优先权(细则4.17(iii))

本国际公布:

— 包括国际检索报告(条约第21条(3))。

feature representation of said image and the plurality of clustering center representations, so as to obtain a category allocation matrix; and performing classification by using the plurality of clustering center representations and the category allocation matrix, so as to obtain a classification result indicating whether said image belongs to a target category. By means of the present application, an image classification result can be more accurate.

(57) 摘要: 本申请实施例公开了一种图像分类方法、训练图像分类模型的方法及装置。主要技术方案包括: 获取待分类的图像; 对所述待分类的图像进行特征提取, 得到所述图像的特征表示; 利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理, 得到多个聚类中心表示; 利用所述图像的特征表示和所述多个聚类中心表示进行解码处理, 得到类别分配矩阵; 利用所述多个聚类中心表示和所述类别分配矩阵进行分类, 得到所述待分类的图像是否属于目标类别的分类结果。通过本申请能够使得图像分类结果具有更高的准确率。

图像分类方法、训练图像分类模型的方法及装置

5 本申请要求于 2023 年 07 月 04 日提交中国专利局、申请号为 202310822213.1、申请名称为“图像分类方法、训练图像分类模型的方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

10 本申请涉及计算机视觉技术领域，特别是涉及一种图像分类的方法、训练图像分类模型的方法及装置。

背景技术

15 图像分类是根据图像的语义信息对不同类别的图像进行区分，是计算机视觉中重要的基础问题。图像分类在很多领域都有着广泛的应用，例如：交通领域的交通场景识别、电商领域的商品图像自动分类，医学领域的图像识别等等。

在一些特殊领域中对于图像分类的准确率和召回率具有较高的要求。虽然目前已经存在利用深度学习模型来对图像进行分类的相关技术，但分类结果的准确率仍有待提高。

发明内容

20 有鉴于此，本申请提供了一种图像分类的方法、训练图像分类模型的方法及装置，以便于提高图像分类结果的准确率。

本申请提供了如下方案：

第一方面，提供了一种图像分类方法，所述方法包括：

获取待分类的图像；

25 对所述待分类的图像进行特征提取，得到所述图像的特征表示；

利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述图像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

30 利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述待分类的图像是否属于目标类别的分类结果。

根据本申请实施例中的一可实现的方式，所述方法还包括：

利用所述类别分配矩阵对所述待分类的图像进行分割，得到预设类别的图像区域，所述预设类别包括所述目标类别。

根据本申请实施例中一可实现的方式，对所述待分类的图像进行特征提取，得到所述图像的特征表示包括：

对所述待分类的图像进行特征提取，得到多种分辨率下的各元素 Token 的特征表示，将最高分辨率下的各 Token 的特征表示作为所述图像的特征表示。

5 根据本申请实施例中一可实现的方式，利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理包括：

10 利用多个聚类中心的初始表示得到查询矩阵，将查询矩阵输入多层串连的转换 Transformer 网络，各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应，各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵，对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理，得到本层 Transformer 网络输出的查询矩阵；

利用最后一层 Transformer 网络输出的查询矩阵得到所述多个聚类中心表示。

根据本申请实施例中一可实现的方式，利用所述多个聚类中心表示和所述类别分配矩阵进行分类包括：

15 对所述多个聚类中心表示进行求平均处理，得到聚类平均表示；

对所述类别分配矩阵进行池化处理，得到聚类池化特征；

将所述聚类平均表示和所述聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到所述待分类的图像是否属于目标类别的分类结果。

第二方面，提供了一种训练图像分类模型的方法，所述方法包括：

20 获取包括多个训练样本的训练数据，所述训练样本包括图像样本以及所述图像样本是否属于目标类别的标签；

25 利用所述训练数据训练图像分类模型，所述图像分类模型包括：特征提取网络、第一解码网络、第二解码网络和分类网络；所述特征提取网络对所述图像样本进行特征提取，得到所述图像样本的特征表示；所述第一解码网络利用所述图像样本的特征表示对多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；所述第二解码网络利用所述图像样本的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；所述分类网络利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述图像样本是否属于目标类别的分类结果；所述训练的目标包括：最小化所述分类结果与对应标签之间的差异。

30 根据本申请实施例中一可实现的方式，所述训练样本还包括对所述图像样本标注的预设类别的区域掩码；所述图像分类模型还包括分割网络；

所述分割网络利用所述类别分配矩阵对所述图像样本进行分割，得到预设类别的图像区域，所述预设类别包括所述目标类别；

所述训练的目标还包括：最小化所述预设类别的图像区域与对应区域掩码之间的差异。

35 根据本申请实施例中一可实现的方式，所述特征提取网络对所述图像样本进行特征提

取，得到所述图像样本的特征表示包括：对所述图像样本进行特征提取，依次得到多种分辨率下的各元素 Token 的特征表示，将最高分辨率下的各 Token 的特征表示作为所述图像样本的特征表示；

所述第一解码网络包括多层串连的 Transformer 网络，所述多个聚类中心的初始表示作为查询矩阵输入第一层 Transformer 网络，各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应，各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵，对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理，得到本层 Transformer 网络输出的查询矩阵；利用最后一层 Transformer 网络输出的查询矩阵得到所述多个聚类中心表示。

根据本申请实施例中一可实现的方式，所述分类网络利用所述多个聚类中心表示和所述类别分配矩阵进行分类包括：

所述分类网络对所述多个聚类中心表示进行求平均处理，得到聚类平均表示；

对所述类别分配矩阵进行池化处理，得到聚类池化表示；

将所述聚类平均表示和所述聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到所述图像样本是否属于目标类别的分类结果。

第三方面，提供了一种图像分类方法，由云端服务器执行，所述方法包括：

获取来自用户终端的待分类的图像；

对所述待分类的图像进行特征提取，得到所述图像的特征表示；

利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述图像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述待分类的图像是否属于目标类别的分类结果；

将所述分类结果返回给所述用户终端。

第四方面，提供了一种癌症计算机辅助诊断方法，其特征在于，所述方法包括：

获取针对目标器官采集的医学影像；

对所述医学影像进行特征提取，得到所述医学影像的特征表示；

利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于目标病变的分类结果。

第五方面提供了一种胃癌计算机辅助诊断方法，所述方法包括：

获取针对胃部采集的医学影像；

对所述医学影像进行特征提取，得到所述医学影像的特征表示；

利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

5 利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于胃癌的分类结果。

10 第六方面，提供了一种癌症的计算机辅助诊断系统，包括存储器、处理器以及存储在所述存储器上并在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序可用于执行一种癌症的计算机辅助诊断方法，所述方法包括：

获取针对目标器官采集的医学影像；

对所述医学影像进行特征提取，得到所述医学影像的特征表示；

利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

15 利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于目标病变的分类结果。

第七方面，提供了一种图像分类装置，所述装置包括：

20 图像获取单元，被配置为获取待分类的图像；

特征提取单元，被配置为对所述待分类的图像进行特征提取，得到所述图像的特征表示；

第一解码单元，被配置为利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

25 第二解码单元，被配置为利用所述图像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

图像分类单元，被配置为利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述待分类的图像是否属于目标类别的分类结果。

第八方面，提供了一种训练图像分类模型的装置，所述装置包括：

30 样本获取单元，被配置为获取包括多个训练样本的训练数据，所述训练样本包括图像样本以及所述图像样本是否属于目标类别的标签；

模型训练单元，被配置为利用所述训练数据训练图像分类模型，所述图像分类模型包括：特征提取网络、第一解码网络、第二解码网络和分类网络；所述特征提取网络对所述图像样本进行特征提取，得到所述图像样本的特征表示；所述第一解码网络利用所述图像
35 样本的特征表示对多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

所述第二解码网络利用所述图像样本的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；所述分类网络利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述图像样本是否属于目标类别的分类结果；所述训练的目标包括：最小化所述分类结果与对应标签之间的差异。

5 第九方面，提供了一种图像分类装置，所述装置包括：

图像获取单元，被配置为获取针对目标器官采集的医学影像；

特征提取单元，被配置为对所述医学影像进行特征提取，得到所述医学影像的特征表示；

10 第一解码单元，被配置为利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

第二解码单元，被配置为利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

15 图像分类单元，被配置为利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于目标病变的分类结果。第十方面，提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现上述第一方面至第五方面中任一项所述的方法的步骤。

第十一方面，提供了一种电子设备，包括：

一个或多个处理器；以及

20 与所述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被所述一个或多个处理器读取执行时，执行上述第一方面至第五方面中任一项所述的方法的步骤。

第十二方面，提供了一种计算机程序，所述计算机程序在计算机中执行时，令计算机执行上述第一方面至第五方面中任一项所述的方法的步骤。

25 第十三方面，提供了一种计算机程序产品，包括计算机程序，所述计算机程序在计算机中执行时，令计算机执行上述第一方面至第五方面中任一项所述的方法的步骤。

根据本申请提供的具体实施例，本申请公开了以下技术效果：

30 1) 本申请通过预先训练得到的多个聚类中心所包含的高水平的语义，与图像的特征表示共同进行解码处理，得到类别分配矩阵，使得类别分配矩阵能够体现与聚类中心关联的图像特征，即将图像特征匹配到相应的聚类中心，进而使得依据聚类中心表示和类别分配矩阵得到的图像分类结果具有更高的准确率和召回率。

2) 本申请能够进一步利用类别分配矩阵对待分类的图像进行分割，得到预设类别的图像区域，从而为图像分类提供可解释性的参考。

35 3) 本申请中通过提取多分辨率的特征，并利用多分辨率的特征多个聚类中心的初始表示进行交叉注意力处理，使得在得到多个聚类中心表示的过程中能够从多尺度感知图像纹理，并将多个聚类中心的特征表示与图像的特征表示进行重新分配，使得分类过程和分割

过程既对局部语义敏感，又具有全局意识，更进一步提高图像分类和图像分割的准确率和召回率。

4) 在训练图像分类模型的过程中，能够利用分割网络对图像样本进行分割得到的预设类别的图像区域与针对图像样本标注的预设类别的区域掩码之间的差异，对图像分类模型进行辅助训练，从而进一步提高图像分类模型的效果和性能。

当然，实施本申请的任一产品并不一定需要同时达到以上所述的所有优点。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为是本申请实施例所适用的系统架构图；

图 2 为本申请实施例提供的图像分类方法的流程图；

图 3 为本申请实施例提供的图像分类模型的原理性示意图；

图 4 为本申请实施例提供的训练图像分类模型的方法流程图；

图 5 为本申请实施例提供的训练图像分类模型的原理性示意图；

图 6 为本申请实施例提供的应用于医学领域的图像分类方法的流程图；

图 7 为本申请实施例提供的图像分类装置的示意性框图；

图 8 为本申请实施例提供的训练图像分类模型的装置示意性框图；

图 9 为本申请实施例提供的电子设备的示意性框图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员所获得的所有其他实施例，都属于本申请保护的范围。

在本发明实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本发明。在本发明实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。

应当理解，本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……”

时”或“响应于确定”或“响应于检测”。类似地，取决于语境，短语“如果确定”或“如果检测（陈述的条件或事件）”可以被解释成为“当确定时”或“响应于确定”或“当检测（陈述的条件或事件）时”或“响应于检测（陈述的条件或事件）”。

传统基于深度神经网络的图像分类方法大多是从图像中提取图像的特征表示后，利用提取的图像的特征表示直接进行分类。但这种方式在一些复杂场景下得到的图像分类结果的准确率和召回率较低，无法满足实际的场景需求。

有鉴于此，本申请提供了一种全新的图像分类思路。为了方便对本申请的理解，首先对本申请所适用的系统架构进行简单描述。图1示出了可以应用本申请实施例的示例性系统架构，如图1中所示，该系统架构包括服务端的模型训练装置和图像分类装置，还可以包括用户终端。

其中，模型训练装置用以在离线阶段进行模型训练。即在获取训练数据后，可以采用本申请实施例提供的方法进行模型训练，得到图像分类模型。

图像分类装置用以在线上利用已经训练得到的图像分类模型，对待分类的图像进行分类，得到图像是否属于目标类别的分类结果。

模型训练装置和图像分类装置可以分别设置为独立的服务器，也可以设置于同一个服务器或服务器群组，还可以设置于独立的或者同一云服务器。云服务器又称为云计算服务器或云主机，是云计算服务体系中的一项主机产品，以解决传统物理主机与虚拟专用服务器（VPs，Virtual Private Server）服务中存在的管理难度大，服务扩展性弱的缺陷。模型训练装置和图像分割装置还可以设置于具有较强计算能力的计算机终端。

作为其中一种可实现的方式，用户可以通过用户终端将待分类的图像通过网络发送给服务端的图像分类装置。由图像分类装置采用本申请实施例提供的方法对待分类的图像进行分类后，将分类结果返回给用户终端。

上述用户终端可以包括但不限于智能移动终端、智能家居设备、可穿戴式设备、智能医疗设备、PC（Personal Computer，个人计算机）等。其中智能移动设备可以包括诸如手机、平板电脑、笔记本电脑、PDA（Personal Digital Assistant，个人数字助理）、互联网汽车等。智能家居设备可以包括诸如智能电视、智能冰箱等等。可穿戴式设备可以包括诸如智能手表、智能眼镜、智能手环、VR（Virtual Reality，虚拟现实）设备、AR（Augmented Reality，增强现实设备）、混合现实设备（即可以支持虚拟现实和增强现实的设备）等等。

需要说明的是，上述图像分类装置除了在线上进行图像分类之外，也可以采用离线的方式进行图像分类，例如针对批量的待分类图像分别进行图像分类。

应该理解，图1中的模型训练装置、图像分类装置、图像分类模型和用户终端的数目仅仅是示意性的。根据实现需要，可以具有任意数目的模型训练装置、图像分类装置、图像分类模型和用户终端。

需要说明的是，本公开中涉及的“第一”、“第二”等限定并不具备大小、顺序和数量等方面的限制，仅仅用以在名称上加以区分。例如“第一图像分割模型”和“第二图像分割模型”

用以在名称上区分两条语音。再例如“第一分割结果”和“第二分割结果”用以在名称上区分两个分割结果。再例如“第一目标”和“第二目标”用以在名称上区分两种目标。等等。

图2为本申请实施例提供的图像分类方法的流程图，该方法可以由图1所示系统中的图像分类装置执行。如图2中所示，该方法可以包括以下步骤：

步骤202：获取待分类图像。

步骤204：对待分类的图像进行特征提取，得到图像的特征表示。

步骤206：利用图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示。

步骤208：利用图像的特征表示和多个聚类中心表示进行解码处理，得到类别分配矩阵。

步骤210：利用多个聚类中心表示和类别分配矩阵进行分类，得到待分类的图像是否属于目标类别的分类结果。

由上述流程可以看出，本申请通过预先训练得到的多个聚类中心所包含的高水平的语义，与图像的特征表示共同进行解码处理，得到类别分配矩阵，使得类别分配矩阵能够体现与聚类中心关联的图像特征，即将图像特征匹配到相应的聚类中心，进而使得依据聚类中心表示和类别分配矩阵得到的图像分类结果具有更高的准确率。

下面对上述各步骤进行详细描述。首先对上述步骤202即“获取待分类的图像”进行详细描述。

在本申请中涉及的待分类的图像可以是二维图像，也可以是三维图像。可以是灰度图像，也可以是彩色图像。

在不同的应用领域待分类的图像可以包含不同的内容。例如，在交通领域，待分类的图像通常为包含诸如车辆、行人、道路、交通设施等交通元素的图像，图像分类的目的可以是确定图像是否属于特定的交通场景。再例如，在电商领域，待分类的图像通常为包含商品信息的图像，图像分类的目的可以是确定图像是否属于特定的商品类目。再例如，在医学领域，图像通常为医学影像，例如CT（Computed Tomography，断层扫描图像）、MRI（Magnetic Resonance Imaging，核磁共振图像）、超音波检查图像等，包含肺、肝脏、胰腺、结肠等器官，图像分类的目的可以是确定图像是否属于特定类别的器官区域，或者是否存在某种特定类别的异常。

上述流程中的步骤204~210可以由预先训练得到的图像分类模型实现。如图3中所示，本申请实施例提供的图像分类模型可以包括：特征提取网络、第一解码网络、第二解码网络和分类网络。

其中，上述步骤204即“对待分类的图像进行特征提取，得到图像的特征表示”可以由上述的特征提取网络执行。

特征提取网络可以基于Transformer（转换）网络实现，实现图像特征的提取，得到待

分类的图像中各 Token（元素）的特征表示。例如特征提取网络可以采用诸如 VIT（Vision Transformer，视觉转换器）、RAN（Residual Attention Network，残差注意力网络）、U-Net（U-Net 是 FCN（Fully Convolutional Network，全卷积神经网络）的一种变体，其提出是为了解决生物医学影像的问题，由于效果很好后来被广泛地应用于图像分割的各个领域）等。

图像的各 Token 指的是构成图像的元素。对于图像而言，将图像切分成不重叠的区块序列，则图像中的区块以及序列的起始符均为 Token。对于二维图像而言，其中的区块可以由一个或多个像素构成。对于三维图像而言，其中的区块可以由一个或多个体素构成。

作为其中一种优选的实施方式，可以采用诸如 U-Net 作为特征提取网络，得到多种分辨率下的各 Token 的特征表示，通过这种多尺度特征来提高图像分类的准确性。具体地，U-Net 采用的是 Encoder（编码器）-Decoder（解码器）架构，Encoder 进行下采样，Decoder 进行上采样和特征拼接，得到多个不同分辨率的特征表示，即先得到低分辨率下各 Token 的特征表示，进而依次得到更高分辨率下各 Token 的特征表示。然后将最高分辨率下的各 Token 的特征表示作为图像的特征表示。

如图 3 中所示，假设输入的待分类图像表示为 X ，通过特征提取网络得到图像的特征表示 F ， $F \in \mathbb{R}^{A \times H \times W \times D}$ 。其中， A 为特征维度， H 、 W 和 D 分别表示三维图像（以输入的待分类图像为三维图像为例）的高、宽和长。

上述步骤 206 即“利用图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示”可以由图像分类模型中的第一解码网络执行。

本步骤中，第一解码网络实际上是将一组聚类中心的初始表示，通过图像的特征表示的交叉注意力，转化为融合了图像语义的聚类中心表示。

若特征提取网络得到多个不同分辨率的特征表示，则作为其中一种优选的实施方式，第一解码网络可以包括多层串连的 Transformer 网络。可以利用多个聚类中心的初始表示得到查询矩阵，将查询矩阵输入多层串连的 Transformer 网络，各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应，各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵，对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理，得到本层 Transformer 网络输出的查询矩阵；利用最后一层 Transformer 网络输出的查询矩阵得到多个聚类中心表示。

每一层 Transformer 网络执行的处理可以表示如下：

$$C_n = C_{n-1} + \arg \max_N (Q(K)^T) V \quad (1)$$

其中， C_n 表示第 n 层 Transformer 网络输出的多个聚类中心表示对应的矩阵， C_{n-1} 为第 $n-1$ 层 Transformer 网络输出的多个聚类中心表示对应的矩阵，对于第一层 Transformer 网络，其输入的 C_{n-1} 为多个聚类中心的初始表示对应的矩阵 $C_{initial}$ 。 Q 由 C_{n-1} 得到， K 和 V 均由该层 Transformer 网络对应分辨率下的各 Token 的特征表示得到。 T 表示转置处理。 N 为聚类中心的数量。

从上述过程中可以看出，第一解码器相当于依据初始聚类中心对图像中的各 Token 进行分组，其中交叉注意力的处理与 K-means 聚类算法类似，使用 argmax 处理替代传统 Transformer 网络中的 softmax 处理。最终得到的多个聚类中心表示对应的矩阵表示为 $C \in \mathbb{R}^{N \times A}$ 。

5 上述步骤 208 即“利用图像的特征表示和多个聚类中心表示进行解码处理，得到类别分配矩阵”可以由上述的第二解码网络执行。

本步骤实际上是基于 N 个聚类中心对各 Token 的特征表示进行重新分配，分配到不同的聚类中，因此将本步骤得到的结果称为类别分配矩阵。

10 第二解码网络可以将多个聚类中心表示对应的矩阵 C 和图像的特征表示 F 进行矩阵相乘，然后进行 Softmax 得到类别分配矩阵对应的矩阵 M ，可以表示为：

$$M = \text{Softmax}_N(CF) \quad (2)$$

其中， $\text{Softmax}_N()$ 表示在 N 聚类中心维度上进行的 Softmax （归一化指数函数）的处理。

15 上述步骤 210 即“利用多个聚类中心表示和类别分配矩阵进行分类，得到待分类的图像是否属于目标类别的分类结果”可以由图 3 所示的图像分类模型中的分类网络执行。

由于学习到的多个聚类中心表示具有高水平的语义，同时体现了聚类之间的差异和聚类内部的相似性，因此，将多个聚类中心表示和类别分配矩阵进行结合考虑进行分类，相比较直接利用图像的特征表示进行分类而言，能够显著提高分类的精确度和召回率。

20 作为其中一种可实现的方式，可以对多个聚类中心表示进行求平均处理，得到聚类平均表示；以及对类别分配矩阵进行池化处理，得到聚类池化表示；然后将聚类平均表示和所述聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到待分类的图像是否属于目标类别的分类结果。

25 例如，聚类平均表示可以在通道维度上对多个聚类中心表示进行求平均处理后得到的。除了求平均处理之外，也可以采用其他处理方式将矩阵 C 转化成向量。在图 3 中，聚类平均表示为 $\bar{C}$ 。

对类别分配矩阵进行池化处理时，可以采用诸如全局最大池化等。在图 3 中，池化处理后得到的聚类池化表示为 $\bar{M}$ 。

30 将聚类平均表示和聚类池化表示进行的整合可以是进行拼接，如图 3 中所示，可以将 $\bar{C}$ 和 $\bar{M}$ 进行拼接后，经过若干层 MLP（Multilayer Perceptron，多层感知机），得到待分类的图像是否属于目标类别的分类结果 $\hat{P}$ 。

在本申请实施例中，分类结果可以是二分类，也可以是多分类。以二分类为例，输出的分类结果可以是属于目标类别，或者不属于目标类别。

更进一步地，在本申请实施例中，还可以由图像分类模型中的分割网络利用类别分配矩阵对待分类的图像进行分割，得到预设类别的图像区域。

35 类别分配矩阵是基于 N 个聚类中心对各 Token 的特征表示进行重新分配，以分配到不

同的聚类中得到的，因此可以将属于同一聚类中的各 Token 作为一个整体来对待。因此，可以将类别分配矩阵投影到 K 个通道，从而得到 K 个类别的图像区域即分割结果 $\hat{Y}$ ， K 为预设的大于 1 的正整数。其中，上述 K 个类别中包括分类网络对应的目标类别。待分类图像的分割结果可以为分类结果提供可解释性，供用户参考并了解目标类别在待分类的图像中的区域。

下面对上述实施例采用的图像分类模型的训练过程进行详细描述。图 4 为本申请实施例提供的训练图像分类模型的方法流程图，该方法流程可以由图 1 所示系统中的模型训练装置执行。如图 4 中所示，该方法可以包括以下步骤：

步骤 402：获取包括多个训练样本的训练数据，训练样本包括图像样本以及图像样本是否属于目标类别的标签。

在本申请实施例中，可以获取一些已知属于目标类别或者不属于目标类别的图像作为图像样本，并对图像样本标注是否属于目标类别的标签。或者，获取一些图像作为图像样本，人工对图像样本标注是否属于目标类别的标签。这些图像样本可以根据实际需求从具体应用领域中获取。

更进一步地，还可以对图像样本标注预设类别的区域掩码 (Mask)，即对图像样本中预设类别的区域进行标注，形成特定类别的区域掩码。可以采用人工标注的方式，也可以获取已知预设类别的区域的图像样本，将其中预设类别的区域进行标注形成区域掩码。上述预设类别包括目标类别。

例如训练数据集 S 可以表示为： $\{(X_i, Y_i, P_i) | i = 1, 2, \dots, m\}$ ，其中， X_i 为图像样本， P_i 可以表示为对 X_i 标注的是否属于目标类别的标签， Y_i 为对 X_i 标注的预设类别的区域掩码， m 为训练样本的数量。

步骤 404：利用训练数据训练图像分类模型，图像分类模型包括：特征提取网络、第一解码网络、第二解码网络和分类网络；特征提取网络对图像样本进行特征提取，得到图像样本的特征表示；第一解码网络利用图像样本的特征表示对多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；第二解码网络利用图像样本的特征表示和多个聚类中心表示进行解码处理，得到类别分配矩阵；分类网络利用多个聚类中心表示和类别分配矩阵进行分类，得到图像样本是否属于目标类别的分类结果；训练的目标包括：最小化分类结果与对应标签之间的差异。

如图 5 中所示，特征提取网络对图像样本 X_i 进行特征提取，得到图像样本的特征表示 F_i 。特征提取网络可以基于 Transformer (转换) 网络实现，实现图像特征的提取，得到图像样本中各 Token (元素) 的特征表示。

作为其中一种可实现的方式特征提取网络可以对图像样本进行特征提取，依次得到多种分辨率下的各元素 Token 的特征表示，将最高分辨率下的各 Token 的特征表示作为图像样本的特征表示。例如可以采用诸如 U-Net 作为特征提取网络，得到多种分辨率下的各 Token 的特征表示，通过这种多尺度特征来提高图像分类的准确性。具体地，U-Net 采用的

是 Encoder (编码器) -Decoder (解码器) 架构, Encoder 进行下采样, Decoder 进行上采样和特征拼接, 得到多个不同分辨率的特征表示, 即先得到低分辨率下各 Token 的特征表示, 进而依次得到更高分辨率下各 Token 的特征表示。然后将最高分辨率下的各 Token 的特征表示作为图像的特征表示。

5 第一解码网络利用图像样本的特征表示 F_i 对多个聚类中心的初始表示 (对应矩阵表示为 $C_{initial}$) 进行交叉注意力处理, 得到多个聚类中心表示 (对应的矩阵表示为 C_i)。

若特征提取网络得到多个不同分辨率的特征表示, 则作为其中一种优选的实施方式, 第一解码网络可以包括多层串连的 Transformer 网络, 多个聚类中心的初始表示作为查询矩阵输入第一层 Transformer 网络, 各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应, 各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵, 对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理, 得到本层 Transformer 网络输出的查询矩阵; 利用最后一层 Transformer 网络输出的查询矩阵得到多个聚类中心表示。

15 第二解码网络利用图像样本的特征表示 F_i 和多个聚类中心表示 C_i 进行解码处理, 得到类别分配矩阵 M_i 。实际上是基于多个聚类中心对各 Token 的特征表示进行重新分配, 分配到不同的聚类中, 因此将第二解码网络得到的结果称为类别分配矩阵。

分类网络利用多个聚类中心表示 C_i 和类别分配矩阵 M_i 进行分类, 得到图像样本是否属于目标类别的分类结果。

20 作为其中一种可实现的方式, 分类网络可以对多个聚类中心表示进行求平均处理, 得到聚类平均表示; 对类别分配矩阵进行池化处理, 得到聚类池化表示; 将聚类平均表示和聚类池化表示进行整合, 利用整合得到的特征表示进行分类, 得到图像样本是否属于目标类别的分类结果。

25 例如, 聚类平均表示可以在通道维度上对多个聚类中心表示进行求平均处理后得到的。除了求平均处理之外, 也可以采用其他处理方式将矩阵 C_i 转化成向量。在图 5 中, 聚类平均表示为 $\bar{C}_i$ 。

对类别分配矩阵进行池化处理时, 可以采用诸如全局最大池化等。在图 5 中, 池化处理后得到的聚类池化表示为 $\bar{M}_i$ 。

30 将聚类平均表示和聚类池化表示进行的整合可以是进行拼接, 如图 5 中所示, 可以将 $\bar{C}_i$ 和 $\bar{M}_i$ 进行拼接后, 经过若干层 MLP (Multilayer Perceptron, 多层感知机), 得到待分类的图像是否属于目标类别的分类结果 $\hat{P}_i$ 。

在本申请实施例中, 分类结果可以是二分类, 也可以是多分类。以二分类为例, 输出的分类结果可以是属于目标类别, 或者不属于目标类别。

上述训练过程中训练的目标包括最小化分类结果 $\hat{P}_i$ 与训练样本中针对 X_i 标注的标签 P_i 之间的差异。

35 更进一步地, 在本申请实施例中, 还可以由图像分类模型中的分割网络利用类别分配

矩阵对图像样本进行分割，得到预设类别的图像区域。

类别分配矩阵是基于多个聚类中心对各 Token 的特征表示进行重新分配，以分配到不同的聚类中得到的，因此可以将属于同一聚类中的各 Token 作为一个整体来对待。因此，可以将类别分配矩阵投影到 K 个通道，从而得到 K 个类别的图像区域即分割结果 $\hat{Y}_i$ ， K 为预设的大于 1 的正整数。其中，上述 K 个类别中包括分类网络对应的目标类别。

上述训练过程中训练的目标可以进一步包括：最小化预设类别的图像区域 $\hat{Y}_i$ 与对应区域掩码 Y_i 之间的差异。该训练的目标可以用以在模型训练过程中辅助对分类结果的学习。

作为其中一种可实现的方式，在本说明书实施例中，可以依据上述训练目标构造损失函数，在每一轮迭代中利用损失函数的取值，采用诸如梯度下降等方式更新模型参数，直至满足预设的训练结束条件。其中训练结束条件可以包括诸如损失函数的取值小于或等于预设的损失函数阈值，迭代次数达到预设的次数阈值等。

若采用上述两个训练的目标，则可以构造总损失函数 L ，例如：

$$L = \frac{1}{b} \sum_{i=1}^b L_{seg}(\hat{Y}_i, Y_i) + L_{cls}(\hat{P}_i, P_i)$$

其中， b 为一个批 (batch) 的训练样本的数量， $L_{seg}(\hat{Y}_i, Y_i)$ 体现 $\hat{Y}_i$ 和 Y_i 之间的差异， $L_{cls}(\hat{P}_i, P_i)$ 体现 $\hat{P}_i$ 和 P_i 之间的差异。

本申请上述实施例提供的图像分类方法可以应用于多种场景。例如，可以将遥感图像作为待分类的图像进行矿物晶体检测、石油探测等，即采用本申请实施例提供的方法对遥感图像进行分类，以确定遥感图像是否属于特定矿物晶体类别，同时可以对遥感图像进行分割，得到各种类别的区域，其中包括特定矿物晶体区域，从而提供可解释性参考。再例如，可以用于一些恶劣环境下的设备异常检测，即可以采用无人机、巡检机器人等设备采集设备图像，将设备图像作为待分类图像，采用本申请实施例提供的方法对设备图像进行分类，以确定是否属于异常类别，即就是否存在异常。同时可以对设备图像进行分割，得到背景区域、设备区域、异常区域等来提供可解释性参考。

还可以应用于其他场景，但其中一个较为重要的场景是医学场景，可以利用本申请实施例提供的方法对医学影像进行分类，以确定是否存在特定病灶。下面以此应用场景进行举例描述。

图 6 为本申请实施例提供的应用于医学场景的图像分类方法流程图，如图 6 中所示，该方法可以具体包括以下步骤：

步骤 602：获取针对目标器官采集的医学影像。

在此采集的医学影像是采用非侵入式的采集方式，例如 CT (Computed Tomography, 断层扫描)、MRI (Magnetic Resonance Imaging, 核磁共振)、超声波检查等。目标器官可以是胃、肺、肝脏、胰腺、结肠等器官。例如，可以是包含胃部的平扫 CT (Non-Contrast CT) 图像等。

以胃部的平扫 CT 图像为例，通常平扫 CT 设备扫描一圈会产生多层图像，因此，平扫 CT 图像 $X \in \mathbb{R}^{H \times W \times D}$ ，其中 H 、 W 和 D 分别表示三维图像 (以输入的待分类图像为三维图

像为例)的高、宽和层数。

步骤 604: 对医学影像进行特征提取, 得到医学影像的特征表示。

步骤 606: 利用医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理, 得到多个聚类中心表示。

5 上述步骤 604 和步骤 606 可以参见之前实施例中针对步骤 204 和步骤 206 的记载, 在此不做赘述。

聚类中心的数量 N 为超参数, N 的取值通常大于后续分割区域的类别数量 K , 因为诸如胃癌等病变可能存在多个亚种, 例如胃癌的不同发展期体现出的特征是不同的, 因此会取更多数量的聚类中心。 N 一般取经验值或实验值, 例如取 8。

10 步骤 608: 利用医学影像的特征表示和多个聚类中心表示进行解码处理, 得到类别分配矩阵。

步骤 610: 利用多个聚类中心表示和类别分配矩阵进行分类, 得到医学影像是否属于目标病变的分类结果。

15 上述步骤 608 和步骤 610 可以参见之前实施例中针对步骤 208 和步骤 210 的记载, 在此不做赘述。

以胃部的平扫 CT 图像为例, 步骤 610 得到的分类结果可以是一个 2 分类, 即正常或者胃癌。从而实现了基于胃部的平扫 CT 图像检测是否有胃癌嫌疑的结果以提供给医师进行提醒和参考。其中, 基于多分辨率的各 Token 的特征表示对多个聚类中心的初始特征表示进行交叉注意力处理, 并将多个聚类中心的特征表示与图像的特征表示进行重新分配, 使得分类过程既能够对局部组织敏感, 又对器官生理学具有全局意识。

20 更进一步地, 在本申请实施例中, 还可以由图像分类模型中的分割网络利用类别分配矩阵对医学影像进行分割, 得到预设类别的图像区域。其中预设类别可以包括背景、器官和病灶。仍以胃部的平扫 CT 图像为例, 可以分割得到背景区域、胃部区域和胃癌区域, 从而提供给医师作为可解释性的参考。

25 在本申请实施例中, 还可以对医学影像进行特征提取, 得到多种分辨率下的各元素 Token 的特征表示, 将最高分辨率下的各 Token 的特征表示作为医学影像的特征表示。

若特征提取网络得到多个不同分辨率的特征表示, 则作为其中一种优选的实施方式, 可以利用多个聚类中心的初始表示得到查询矩阵, 将查询矩阵输入多层串连的转换 Transformer 网络, 各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应, 各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵, 对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理, 得到本层 Transformer 网络输出的查询矩阵; 利用最后一层 Transformer 网络输出的查询矩阵得到多个聚类中心表示。

30 作为其中一种可实现的方式, 可以对多个聚类中心表示进行求平均处理, 得到聚类平均表示; 以及对类别分配矩阵进行池化处理, 得到聚类池化表示; 然后将聚类平均表示和所述聚类池化表示进行整合, 利用整合得到的特征表示进行分类, 得到医学图像是否属于

目标病变的分类结果。本申请实施例还提供一种胃癌计算机辅助诊断方法，所述方法包括：

获取针对胃部采集的医学影像；

对所述医学影像进行特征提取，得到所述医学影像的特征表示；

利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉

5 注意力处理，得到多个聚类中心表示；

利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于胃癌的分类结果。

10 在采用图 4 所示流程对检测胃癌的图像分类模型进行训练时，图像样本可以是已知正常或胃癌的胃部的平扫 CT 图像作为各 X_i ，对各平扫 CT 图像标注正常或胃癌的标签作为 P_i 。由于平扫 CT 的区域标注较为困难，因此，可以由医师对各 X_i 对应的增强 CT (Contrast CT) 图像进行背景区域、胃部区域和胃癌区域的标注后，将标注后的增强 CT 图像与平扫 CT 图像对齐，从而形成相对粗糙但可靠性高的各区域掩码。

15 胃癌是全球范围内癌症相关死亡的第三大原因，5 年生存率约为 33%，如果能够在早期检测出相关症状则可以显著提高 5 年生存率。由于早期胃肿瘤可能只侵犯黏膜和肌层，如果没有进行胃部造影剂注射很难识别。目前诸如钡餐胃部造影、内镜检查、血清胃蛋白酶原水平检查等已有的检测方式都是侵入式的，成本高且副作用大，难以很好地应用于早期胃癌检测。采用本申请实施例提供的上述方式，能够利用诸如胃部平扫 CT 图像等无创、
20 成本低的方式来获取图像，并利用本申请实施例提供的方式，由计算机设备进行图像分类来实现胃癌检测，即最终输出的分类结果为是否属于胃癌，例如输出正常、胃癌的二分类结果，从而作为中间数据给医师或患者提供参考依据或提醒，以方便后续可以采用更进一步地检查和诊断。同时，能够对胃部平扫 CT 图像进行图像分割，分割得到背景区域、胃部区域以及病灶区域（即胃癌所在区域），从而作为中间数据给医师或患者提供可解释性地
25 参考。显然，这是一种无创、低成本且易于推广的新型检测方法，并且在精度上也有很好的效果。

经过实验，本申请实施例提供的方法在 100 例胃癌患者和 148 例正常患者的胃部平扫 CT 图像的测试集上，分类得到胃癌的分类结果的灵敏度为 85.0%，特异性为 92.6%。而放射科医生人为对胃部平扫 CT 图像进行胃癌判别的平均灵敏度为 73.5%，特异性为 84.3%，
30 显然通过本申请实施例提供的基于计算机视觉技术的图像分类方法效果更优。其中灵敏度又称为真阳性率，是指对于胃癌样本进行测试的分类结果为胃癌的概率。特异性又称为真阴性率，是指对于正常个样本进行测试的分类结果为正常的概率。

上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且
35 且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要示出的特定顺序或者

连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

参见下述表 1，表 1 输出了本公开的图像分类方法与三个基准的比较分析。前两个基准使用的两种方法属于“分类分割”（S4C），使用 nnUNet 和 TransUNet。其中，†表示 DeLong 测试（本公开与 nnUNet-S4C）的 $p < 0.05$ ；*表示排列检验（本公开与 nnUNet-S4C 和放射科医生）的 $p < 0.05$ 。如果分割的肿瘤体积超过使验证集的灵敏度和特异性之和最大化的阈值，则病例被分类为阳性。第三条基线（表示为“nnUNet-Joint”）将 CNN 分类头集成到 UNet，并进行端到端训练。可以从测试数据集的 1000 个 bootstrap 副本中获得了 AUC、灵敏度和特异性值的 95%置信区间，以进行统计分析。为了获得统计显著性，可以在两个 AUC（本公开的图像分类方法与比较方法）之间进行 DeLong 检验，并在两个灵敏度或特异性（本公开的图像分类方法与比较方法和放射科医生）之间进行排列检验。

表 1

方法	内部表现 (internal hold-out) (n = 248)			外部 (n = 903)
	AUC	灵敏度(%)	特异性(%)	特异性(%)
放射科医生	-	73.5	84.1	-
nnUNet-S4C	0.907 (0.862, 0.942)	80 (72.0, 87.5)	88.5 (83.3, 93.5)	96.6 (95.2, 97.8)
TransUNet-S4C	0.916 (0.876, 0.952)	82 (74.7, 89.5)	90.5 (86.1, 94.8)	96 (94.8, 97.2)
nnUNet-Joint	0.924 (0.885, 0.959)	81 (73.0, 87.9)	90.5 (85.1, 95.0)	97.6 (96.5, 98.6)
本公开的图像分类方法	0.939† (0.910, 0.964)	85.0* (78.1, 91.1)	92.6* (88.0, 96.5)	97.7 (96.7, 98.7)

参照表 1 所示，本公开的图像分类方法在所有指标上都优于其他三个基准（表 1），特别是在 AUC 和灵敏度方面。本公开的方法的优点是，它凭借掩模 Transformer 的独特架构同时捕获局部和全局信息。它还从集群表示中提取高级语义，使其适合分类并促进整体决策过程。此外，本公开的方法在外部测试集上达到了 97.7%的相当大的特异性，这对于减少在机会性筛选方面的误报和不必要的人力工作量的至关重要。

本公开提出的模型的 ROC 曲线优于由两名经验丰富的放射科医生进行检测的结果。该模型在检测胃癌方面达到了 85.0%的灵敏度，显著超过了医生的平均表现（73.5%），也超过了表现最好的医生（75.0%），同时保持高特异性。

参照表 2 所示，按肿瘤（T）分期分层的患者水平检测和肿瘤水平定位的表现。表 2 将本公开的模型的性能与两位放射科医生的性能进行比较。结果表明，本公开的模型在检测早期肿瘤（T1、T2）方面表现更好，并提供更精确的肿瘤定位。具体来说，本公开的模型检测出 60.0% (6/10) T1 癌症和 77.8% (7/9) T2 癌症，超过了表现最好的专家（50% T1, 55.6% T2）。同时，本公开的模型对 T3 和 T4 肿瘤保持了可靠的检测率和可信

的定位精度（34 个 T3 肿瘤中有 2 个漏检）。

表 2 不同 T 阶段胃癌的患者层面检测和肿瘤层面定位结果(%)

方法	标准	T1	T2	T3	T4	T 阶段信息缺失
nnUNet-S4C	病人	30.0(3/10)	66.7(6/9)	94.1(32/34)	100.0(9/9)	86.1(31/36)
	肿瘤	20.0(2/10)	55.6(5/9)	94.1(32/34)	100.0(9/9)	80.6(29/36)
本公开	病人	60.0(6/10)	77.8(7/9)	94.1(32/34)	100.0(9/9)	86.1(31/36)
	肿瘤	30.0(3/10)	66.7(6/9)	94.1(32/34)	100.0(9/9)	80.6(30/36)
放射科医生 1	病人	50.0(5/10)	55.6(5/9)	76.5(26/34)	88.9(8/9)	77.8(28/36)
放射科医生 2	病人	30.0(3/10)	55.6(5/9)	85.3(29/34)	100.0(9/9)	80.6(29/36)

如表 3 所示，在相对较大的测试患者规模（通过整合内部和外部测试集，n = 1151）的情况下，本公开的方法在胃癌检测的敏感性方面超越或与现有的筛查工具相当，具有相似的特异性水平，如表 3 所示。其中，*表示在测试集中将两个原位肿瘤移除。†表示该比较中只考虑早期胃癌病例，包括原位肿瘤、T1 和 T2 期，其中 19 例中本公开的方法成功检出 17 例。

表 3 在大规模人群中将最先进的血液检测应用在胃癌检测方面、UGIS 和内窥镜筛查的表现，以及高级放射科医生在窄带成像上使用放大胃镜（magnifying endoscopy, ME-NBI）的早期胃癌检出率比较

方法	规格(%)	灵敏度(%)	本公开在相同特异性下的灵敏度(%)
血液检测	99.5	66.7	69.4*
上消化道造影	96.1	36.7	85.0
内窥镜检查	96.0	69.0	85.0
ME-NBI (早期)	74.2	76.7	89.5†

根据另一方面的实施例，提供了一种图像分类装置。图 7 示出根据一个实施例的图像分类装置的示意性框图。如图 7 所示，该装置 700 包括：图像获取单元 701、特征提取单元 702、第一解码单元 703、第二解码单元 704 和图像分类单元 705，还可以进一步图像分割单元 706。其中各组成单元的主要功能如下：

图像获取单元 701，被配置为获取待分类的图像。

特征提取单元 702，被配置为对待分类的图像进行特征提取，得到图像的特征表示。

第一解码单元 703，被配置为利用图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示。

第二解码单元 704，被配置为利用图像的特征表示和多个聚类中心表示进行解码处理，

得到类别分配矩阵。

图像分类单元 705，被配置为利用多个聚类中心表示和类别分配矩阵进行分类，得到待分类的图像是否属于目标类别的分类结果。

更进一步地，图像分割单元 706，被配置为利用类别分配矩阵对待分类的图像进行分割，得到预设类别的图像区域，预设类别包括目标类别。

作为其中一种可实现的方式，上述特征提取单元 702 可以具体被配置为：对待分类的图像进行特征提取，得到多种分辨率下的各元素 Token 的特征表示，将最高分辨率下的各 Token 的特征表示作为图像的特征表示。

相应地，第一解码单元 703 可以利用多个聚类中心的初始表示得到查询矩阵，将查询矩阵输入多层串连的转换 Transformer 网络，各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应，各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵，对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理，得到本层 Transformer 网络输出的查询矩阵；利用最后一层 Transformer 网络输出的查询矩阵得到多个聚类中心表示。

作为其中一种可实现的方式，图像分类单元 705，可以具体被配置为对多个聚类中心表示进行求平均处理，得到聚类平均表示；对类别分配矩阵进行池化处理，得到聚类池化特征；将聚类平均表示和聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到待分类的图像是否属于目标类别的分类结果。

上述特征提取单元 702、第一解码单元 703、第二解码单元 704、图像分类单元 705 和图像分割单元 706 分别对应于图 3 所示图像分类模型中的特征提取网络、第一解码网络、第二解码网络、分类网络和分割网络，具体细节可以参见上述方法实施例中针对图 3 的相关记载，在此不做赘述。

图 8 为本申请实施例提供的训练图像分类模型的装置的示意性框图，如图 8 中所示，该装置可以包括：样本获取单元 801 和模型训练单元 802。其中各组成单元的主要功能如下：

样本获取单元 801，被配置为获取包括多个训练样本的训练数据，训练样本包括图像样本以及图像样本是否属于目标类别的标签。

模型训练单元 802，被配置为利用训练数据训练图像分类模型，图像分类模型包括：特征提取网络、第一解码网络、第二解码网络和分类网络；特征提取网络对图像样本进行特征提取，得到图像样本的特征表示；第一解码网络利用图像样本的特征表示对多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；第二解码网络利用图像样本的特征表示和多个聚类中心表示进行解码处理，得到类别分配矩阵；分类网络利用多个聚类中心表示和类别分配矩阵进行分类，得到图像样本是否属于目标类别的分类结果；训练的目标包括：最小化分类结果与对应标签之间的差异。

作为其中一种可实现的方式，上述训练样本还可以包括对图像样本标注的预设类别的

区域掩码。图像分类模型还包括可以分割网络。

分割网络利用类别分配矩阵对图像样本进行分割，得到预设类别的图像区域，预设类别包括目标类别。相应地，上述训练的目标还可以包括：最小化预设类别的图像区域与对应区域掩码之间的差异。

5 作为其中一种可实现的方式，特征提取网络可以对图像样本进行特征提取，依次得到多种分辨率下的各元素 Token 的特征表示，将最高分辨率下的各 Token 的特征表示作为图像样本的特征表示。

10 相应地，第一解码网络可以包括多层串连的 Transformer 网络，多个聚类中心的初始表示作为查询矩阵输入第一层 Transformer 网络，各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应，各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵，对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理，得到本层 Transformer 网络输出的查询矩阵；利用最后一层 Transformer 网络输出的查询矩阵得到多个聚类中心表示。

15 作为其中一种可实现的方式，分类网络可以对多个聚类中心表示进行求平均处理，得到聚类平均表示；对类别分配矩阵进行池化处理，得到聚类池化表示；将聚类平均表示和聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到图像样本是否属于目标类别的分类结果。

20 根据另一方面的实施例，提供了一种图像分类装置，应用于医学诊断。该装置包括：图像获取单元、特征提取单元、第一解码单元、第二解码单元和图像分类单元，还可以进一步图像分割单元。其中各组成单元的主要功能如下：

图像获取单元，被配置为获取针对目标器官采集的医学影像；

特征提取单元，被配置为对所述医学影像进行特征提取，得到所述医学影像的特征表示；

25 第一解码单元，被配置为利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

第二解码单元，被配置为利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

图像分类单元，被配置为利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于目标病变的分类结果。

30 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是
35 物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需

要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

需要说明的是，本申请所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

另外，本申请实施例还提供了一种癌症计算机辅助诊断系统，包括存储器、处理器以及存储在所述存储器上并在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序可用于执行一种图像分类方法，所述方法包括：

获取针对目标器官采集的医学影像；

对所述医学影像进行特征提取，得到所述医学影像的特征表示；

利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于目标病变的分类结果。

另外，本申请实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现前述方法实施例中任一项所述的方法的步骤。

以及一种电子设备，包括：

一个或多个处理器；以及

与所述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被所述一个或多个处理器读取执行时，执行前述方法实施例中任一项所述的方法的步骤。

本申请还提供了一种计算机程序产品，包括计算机程序，该计算机程序在被处理器执行时实现前述方法实施例中任一项所述的方法的步骤。

本申请还提供了一种计算机程序，其中，当所述计算机程序在计算机中执行时，令计算机执行前述方法实施例中任一项所述的方法的步骤。

其中，图9示例性的展示出了电子设备的架构，具体可以包括处理器910，视频显示适配器911，磁盘驱动器912，输入/输出接口913，网络接口914，以及存储器920。上述处理器910、视频显示适配器911、磁盘驱动器912、输入/输出接口913、网络接口914，与存储器920之间可以通过通信总线930进行通信连接。

其中，处理器910可以采用通用的CPU、微处理器、应用专用集成电路（Application Specific Integrated Circuit，ASIC）、或者一个或多个集成电路等方式实现，用于执行相关程序，以实现本申请所提供的技术方案。

存储器 920 可以采用 ROM (Read Only Memory, 只读存储器)、RAM (Random Access Memory, 随机存取存储器)、静态存储设备, 动态存储设备等形式实现。存储器 920 可以存储用于控制电子设备 900 运行的操作系统 921, 用于控制电子设备 900 的低级别操作的基本输入输出系统(BIOS) 922。另外, 还可以存储网页浏览器 923, 数据存储管理系统 924, 以及图像分类装置/模型训练装置 925 等等。上述图像分类装置/模型训练装置 925 就可以是本申请实施例中具体实现前述各步骤操作的应用程序。总之, 在通过软件或者固件来实现本申请所提供的技术方案时, 相关的程序代码保存在存储器 920 中, 并由处理器 910 来调用执行。

输入/输出接口 913 用于连接输入/输出模块, 以实现信息输入及输出。输入输出/模块可以作为组件配置在设备中(图中未示出), 也可以外接于设备以提供相应功能。其中输入设备可以包括键盘、鼠标、触摸屏、麦克风、各类传感器等, 输出设备可以包括显示器、扬声器、振动器、指示灯等。

网络接口 914 用于连接通信模块(图中未示出), 以实现本设备与其他设备的通信交互。其中通信模块可以通过有线方式(例如 USB、网线等)实现通信, 也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信。

总线 930 包括一通路, 在设备的各个组件(例如处理器 910、视频显示适配器 911、磁盘驱动器 912、输入/输出接口 913、网络接口 914, 与存储器 920)之间传输信息。

需要说明的是, 尽管上述设备仅示出了处理器 910、视频显示适配器 911、磁盘驱动器 912、输入/输出接口 913、网络接口 914, 存储器 920, 总线 930 等, 但是在具体实施过程中, 该设备还可以包括实现正常运行所必需的其他组件。此外, 本领域的技术人员可以理解的是, 上述设备中也可以仅包含实现本申请方案所必需的组件, 而不必包含图中所示的全部组件。

通过以上的实施方式的描述可知, 本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解, 本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以计算机程序产品的形式体现出来, 该计算机程序产品可以存储在存储介质中, 如 ROM/RAM、磁碟、光盘等, 包括若干指令用以使得一台计算机设备(可以是个人计算机, 服务器, 或者网络设备)执行本申请各个实施例或者实施例的某些部分所述的方法。

以上对本申请所提供的技术方案进行了详细介绍, 本文中应用了具体个例对本申请的原理及实施方式进行了阐述, 以上实施例的说明只是用于帮助理解本申请的方法及其核心思想; 同时, 对于本领域的一般技术人员, 依据本申请的思想, 在具体实施方式及应用范围上均会有改变之处。综上所述, 本说明书内容不应理解为对本申请的限制。

权 利 要 求 书

1. 一种图像分类方法，其特征在于，所述方法包括：

获取待分类的图像；

对所述待分类的图像进行特征提取，得到所述图像的特征表示；

5 利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述图像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述待分类的图像是否属于目标类别的分类结果。

10 2. 根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

利用所述类别分配矩阵对所述待分类的图像进行分割，得到预设类别的图像区域，所述预设类别包括所述目标类别。

3. 根据权利要求 1 或 2 所述的方法，其特征在于，对所述待分类的图像进行特征提取，得到所述图像的特征表示包括：

15 对所述待分类的图像进行特征提取，得到多种分辨率下的各元素 Token 的特征表示，将最高分辨率下的各 Token 的特征表示作为所述图像的特征表示。

4. 根据权利要求 3 所述的方法，其特征在于，利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理包括：

20 利用多个聚类中心的初始表示得到查询矩阵，将查询矩阵输入多层串连的转换 Transformer 网络，各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应，各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵，对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理，得到本层 Transformer 网络输出的查询矩阵；

利用最后一层 Transformer 网络输出的查询矩阵得到所述多个聚类中心表示。

25 5. 根据权利要求 1 至 4 中任一项所述的方法，其特征在于，利用所述多个聚类中心表示和所述类别分配矩阵进行分类包括：

对所述多个聚类中心表示进行求平均处理，得到聚类平均表示；

对所述类别分配矩阵进行池化处理，得到聚类池化特征；

30 将所述聚类平均表示和所述聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到所述待分类的图像是否属于目标类别的分类结果。

6. 一种训练图像分类模型的方法，其特征在于，所述方法包括：

获取包括多个训练样本的训练数据，所述训练样本包括图像样本以及所述图像样本是否属于目标类别的标签；

35 利用所述训练数据训练图像分类模型，所述图像分类模型包括：特征提取网络、第一解码网络、第二解码网络和分类网络；所述特征提取网络对所述图像样本进行特征提取，

得到所述图像样本的特征表示；所述第一解码网络利用所述图像样本的特征表示对多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；所述第二解码网络利用所述图像样本的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；所述分类网络利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述图像样本是否属于目标类别的分类结果；所述训练的目标包括：最小化所述分类结果与对应标签之间的差异。

7. 根据权利要求6所述的方法，其特征在于，所述训练样本还包括对所述图像样本标注的预设类别的区域掩码；所述图像分类模型还包括分割网络；

所述分割网络利用所述类别分配矩阵对所述图像样本进行分割，得到预设类别的图像区域，所述预设类别包括所述目标类别；

所述训练的目标还包括：最小化所述预设类别的图像区域与对应区域掩码之间的差异。

8. 根据权利要求6或7所述的方法，其特征在于，所述特征提取网络对所述图像样本进行特征提取，得到所述图像样本的特征表示包括：对所述图像样本进行特征提取，依次得到多种分辨率下的各元素Token的特征表示，将最高分辨率下的各Token的特征表示作为所述图像样本的特征表示；

所述第一解码网络包括多层串连的Transformer网络，所述多个聚类中心的初始表示作为查询矩阵输入第一层Transformer网络，各层Transformer网络按照分辨率从低到高的顺序与各分辨率一一对应，各层Transformer网络利用对应分辨率下的各Token的特征表示得到键矩阵和值矩阵，对输入本层Transformer网络的查询矩阵进行交叉注意力处理，得到本层Transformer网络输出的查询矩阵；利用最后一层Transformer网络输出的查询矩阵得到所述多个聚类中心表示。

9. 根据权利要求6至8中任一项所述的方法，其特征在于，所述分类网络利用所述多个聚类中心表示和所述类别分配矩阵进行分类包括：

所述分类网络对所述多个聚类中心表示进行求平均处理，得到聚类平均表示；

对所述类别分配矩阵进行池化处理，得到聚类池化表示；

将所述聚类平均表示和所述聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到所述图像样本是否属于目标类别的分类结果。

10. 一种图像分类方法，由云端服务器执行，其特征在于，所述方法包括：

获取来自用户终端的待分类的图像；

对所述待分类的图像进行特征提取，得到所述图像的特征表示；

利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述图像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述待分类的图像是否属于目标类别的分类结果；

将所述分类结果返回给所述用户终端。

11. 一种癌症计算机辅助诊断方法，其特征在于，所述方法包括：

获取针对目标器官采集的医学影像；

对所述医学影像进行特征提取，得到所述医学影像的特征表示；

5 利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

10 利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于目标病变的分类结果。

12. 根据权利要求 11 所述的方法，其特征在于，所述方法还包括：

利用所述类别分配矩阵对所述医学影像进行分割，得到预设类别的图像区域，所述预设类别包括所述目标病变。

13. 根据权利要求 11 或 12 所述的方法，其特征在于，对所述医学影像进行特征提取，
15 得到所述医学影像的特征表示包括：

对所述医学影像进行特征提取，得到多种分辨率下的各元素 Token 的特征表示，将最高分辨率下的各 Token 的特征表示作为所述医学影像的特征表示。

14. 根据权利要求 13 所述的方法，其特征在于，利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理包括：

20 利用多个聚类中心的初始表示得到查询矩阵，将查询矩阵输入多层串连的转换 Transformer 网络，各层 Transformer 网络按照分辨率从低到高的顺序与各分辨率一一对应，各层 Transformer 网络利用对应分辨率下的各 Token 的特征表示得到键矩阵和值矩阵，对输入本层 Transformer 网络的查询矩阵进行交叉注意力处理，得到本层 Transformer 网络输出的查询矩阵；

25 利用最后一层 Transformer 网络输出的查询矩阵得到所述多个聚类中心表示。

15. 根据权利要求 11 至 14 中任一项所述的方法，其特征在于，利用所述多个聚类中心表示和所述类别分配矩阵进行分类包括：

对所述多个聚类中心表示进行求平均处理，得到聚类平均表示；

对所述类别分配矩阵进行池化处理，得到聚类池化特征；

30 将所述聚类平均表示和所述聚类池化表示进行整合，利用整合得到的特征表示进行分类，得到所述医学影像是否属于目标病变的分类结果。

16. 一种胃癌计算机辅助诊断方法，所述方法包括：

获取针对胃部采集的医学影像；

对所述医学影像进行特征提取，得到所述医学影像的特征表示；

35 利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉

注意力处理，得到多个聚类中心表示；

利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

5 利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于胃癌的分类结果。

17. 一种癌症计算机辅助诊断系统，包括存储器、处理器以及存储在所述存储器上并在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序可用于执行一种癌症计算机方法，所述方法包括：

获取针对目标器官采集的医学影像；

10 对所述医学影像进行特征提取，得到所述医学影像的特征表示；

利用所述医学影像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述医学影像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

15 利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述医学影像是否属于目标病变的分类结果。

18. 一种图像分类装置，其特征在于，所述装置包括：

图像获取单元，被配置为获取待分类的图像；

20 特征提取单元，被配置为对所述待分类的图像进行特征提取，得到所述图像的特征表示；

第一解码单元，被配置为利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

第二解码单元，被配置为利用所述图像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

25 图像分类单元，被配置为利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述待分类的图像是否属于目标类别的分类结果。

19. 一种训练图像分类模型的装置，其特征在于，所述装置包括：

样本获取单元，被配置为获取包括多个训练样本的训练数据，所述训练样本包括图像样本以及所述图像样本是否属于目标类别的标签；

30 模型训练单元，被配置为利用所述训练数据训练图像分类模型，所述图像分类模型包括：特征提取网络、第一解码网络、第二解码网络和分类网络；所述特征提取网络对所述图像样本进行特征提取，得到所述图像样本的特征表示；所述第一解码网络利用所述图像样本的特征表示对多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；所述第二解码网络利用所述图像样本的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；所述分类网络利用所述多个聚类中心表示和所述类别分配矩阵进行分

35

类，得到所述图像样本是否属于目标类别的分类结果；所述训练的目标包括：最小化所述分类结果与对应标签之间的差异。

20. 一种图像分类系统，包括存储器、处理器以及存储在所述存储器上并在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序可用于执行一种图像分类方法，所述方法包括：

获取待分类的图像；

对所述待分类的图像进行特征提取，得到所述图像的特征表示；

利用所述图像的特征表示对预先训练得到的多个聚类中心的初始表示进行交叉注意力处理，得到多个聚类中心表示；

利用所述图像的特征表示和所述多个聚类中心表示进行解码处理，得到类别分配矩阵；

利用所述多个聚类中心表示和所述类别分配矩阵进行分类，得到所述待分类的图像是否属于目标类别的分类结果。

21. 一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现权利要求 1 至 16 中任一项所述的方法的步骤。

22. 一种电子设备，其特征在于，包括：

一个或多个处理器；以及

与所述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被所述一个或多个处理器读取执行时，执行权利要求 1 至 16 中任一项所述的方法的步骤。

23. 一种计算机程序产品，包括计算机可执行指令，该计算机可执行指令被处理器执行时实现权利要求 1 至 16 任一项所述方法的步骤。

24. 一种计算机程序，该计算机程序被处理器执行时实现权利要求 1 至 16 任一项所述方法的步骤。

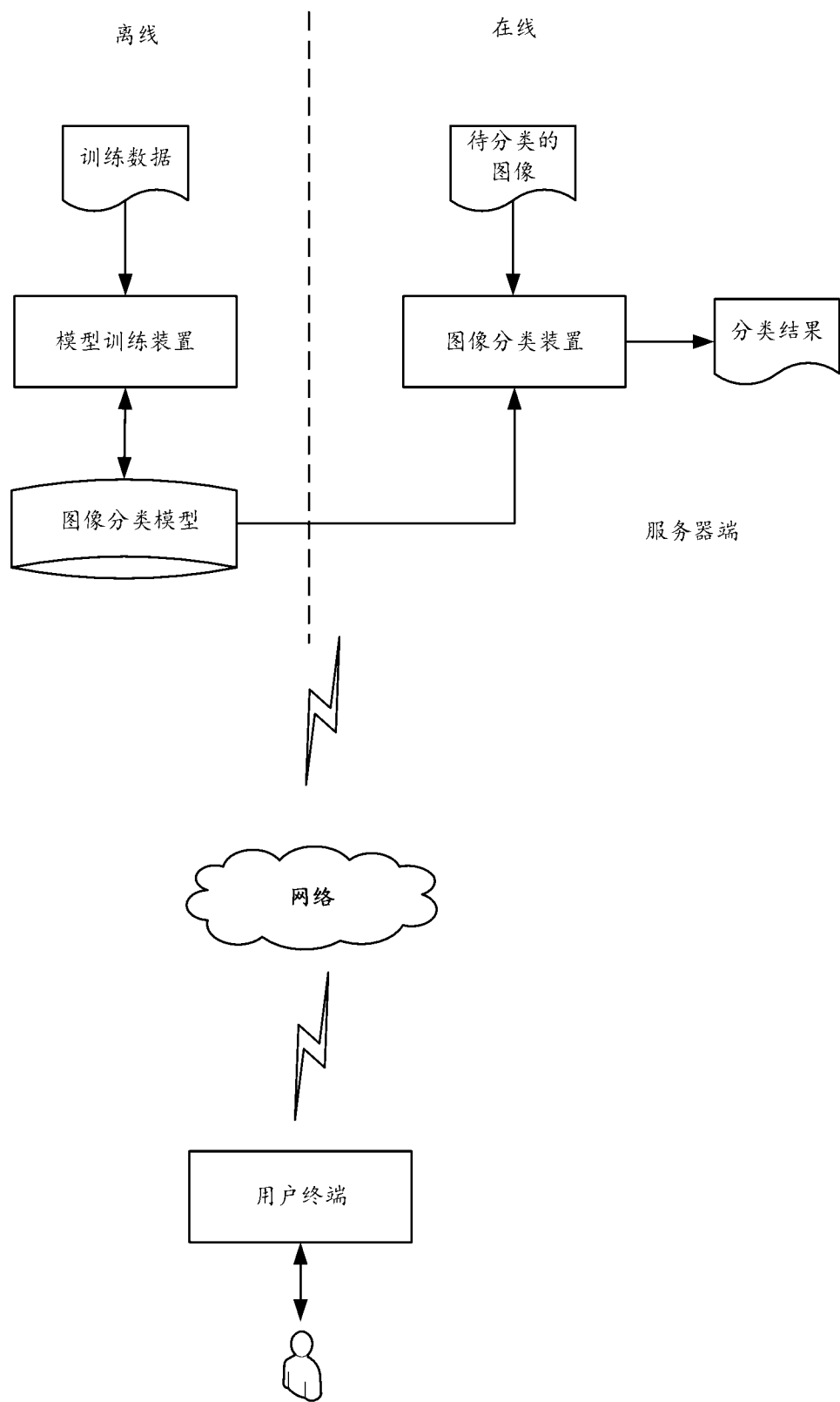


图 1

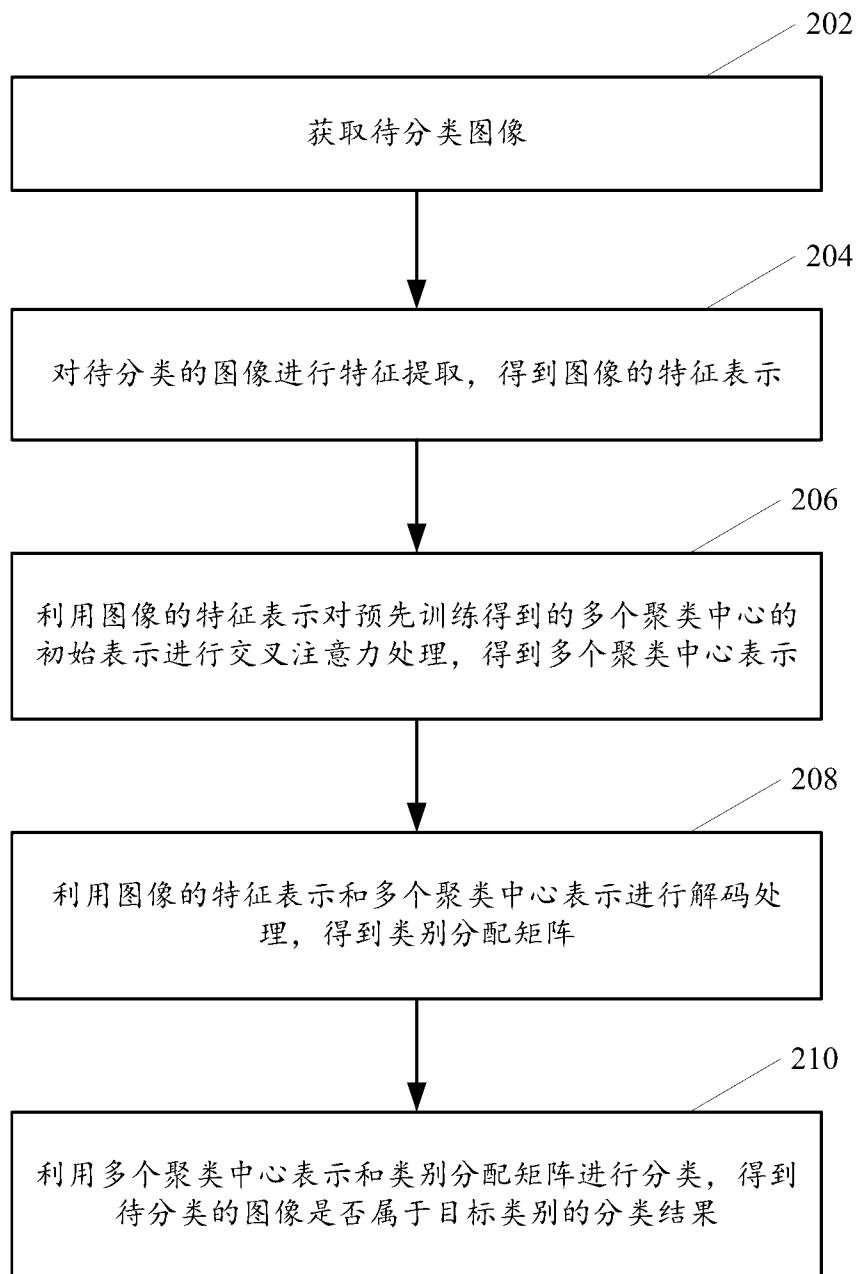


图 2

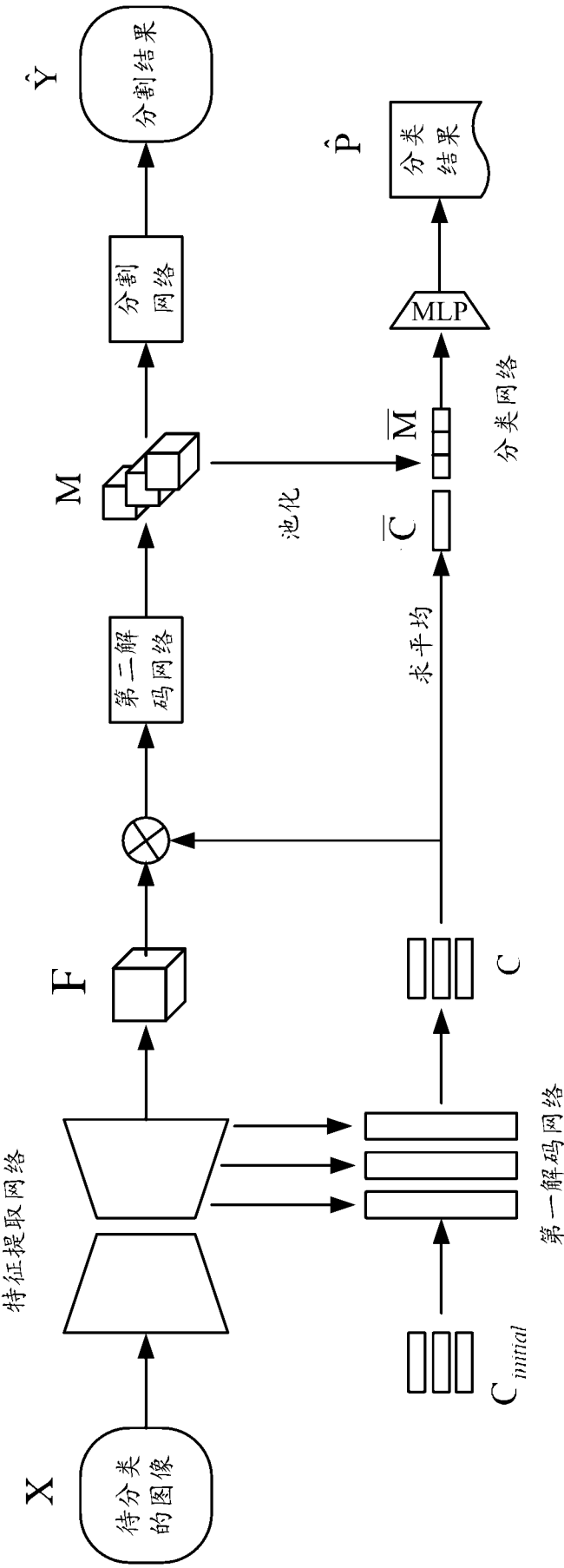


图3

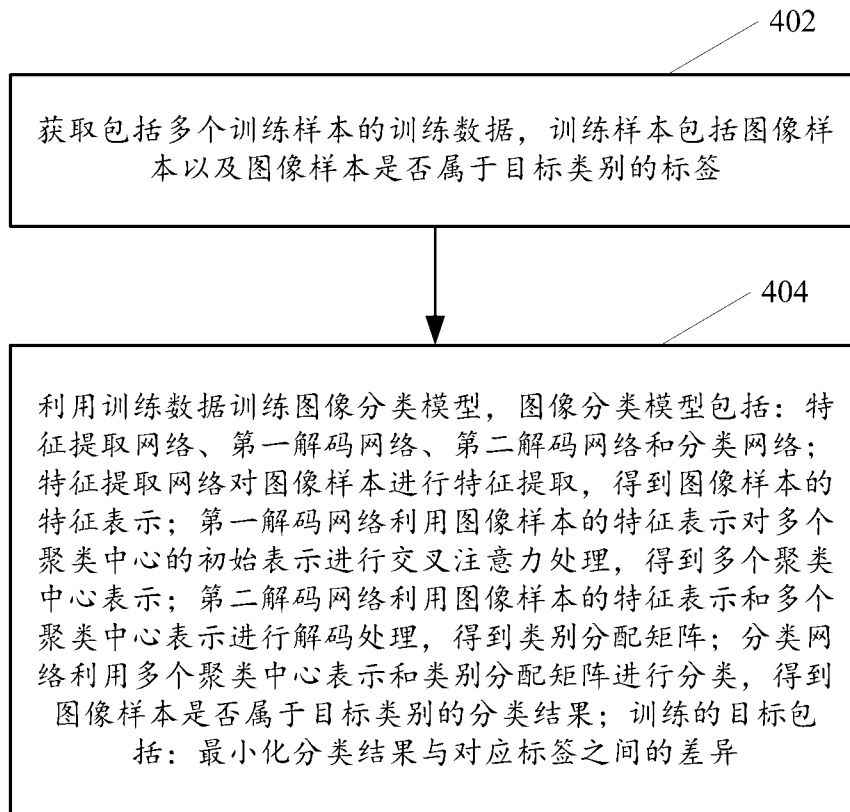


图 4

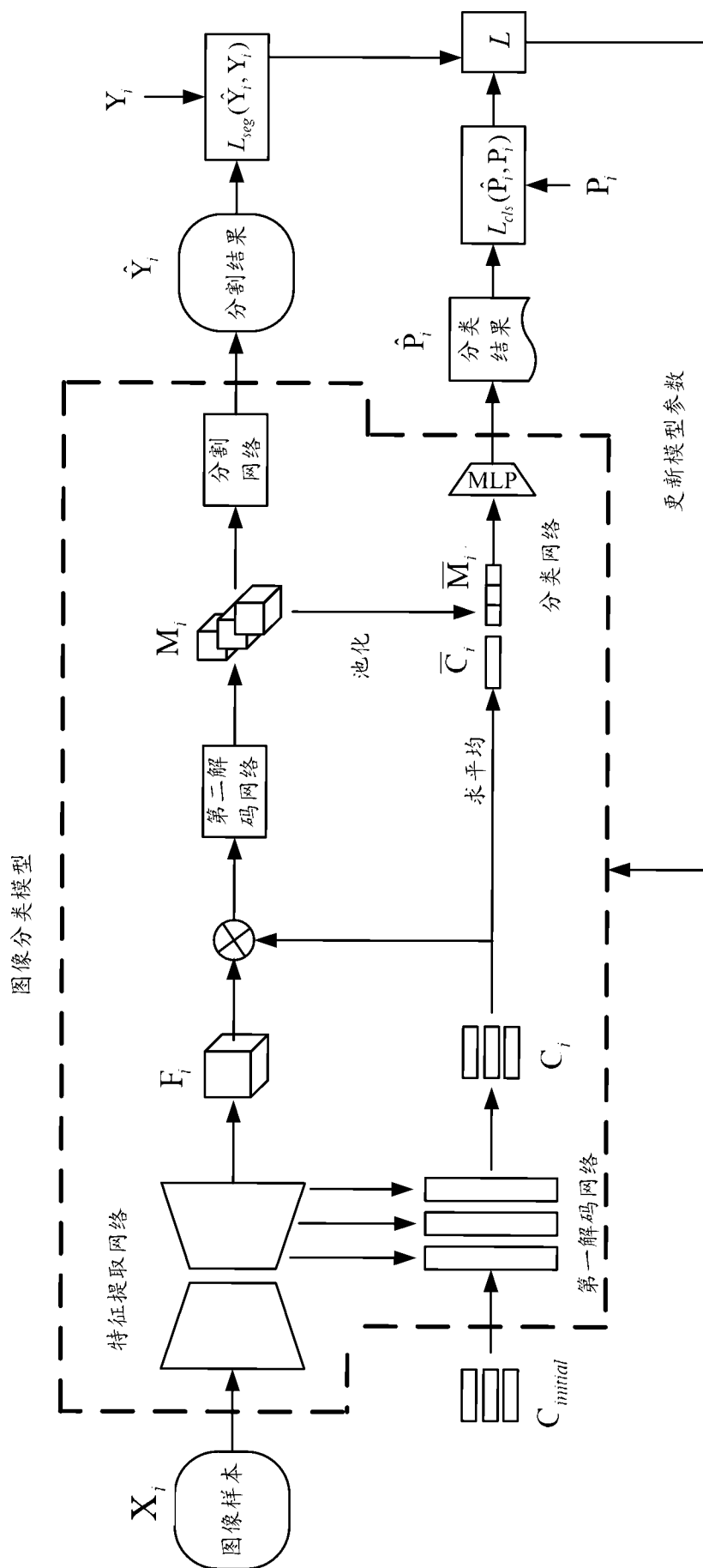


图5

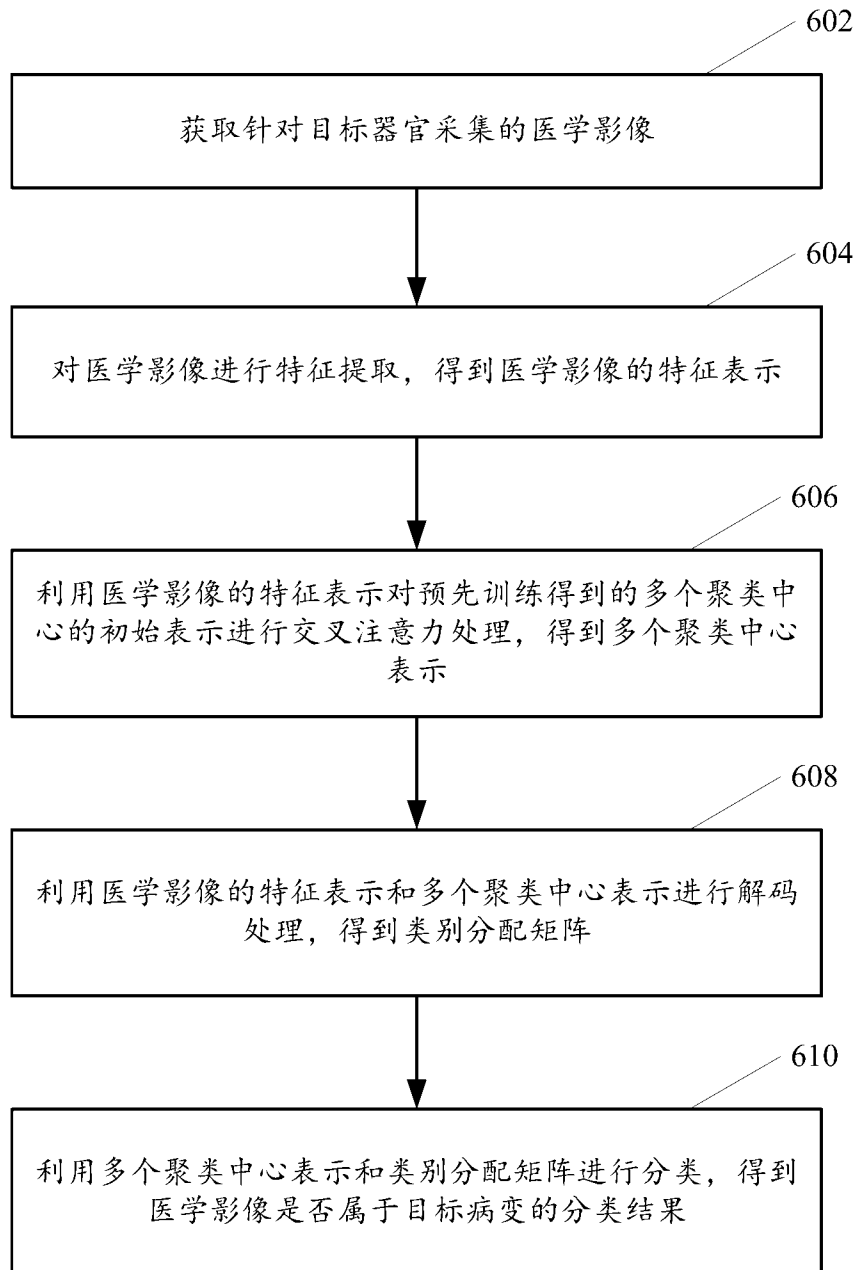


图 6

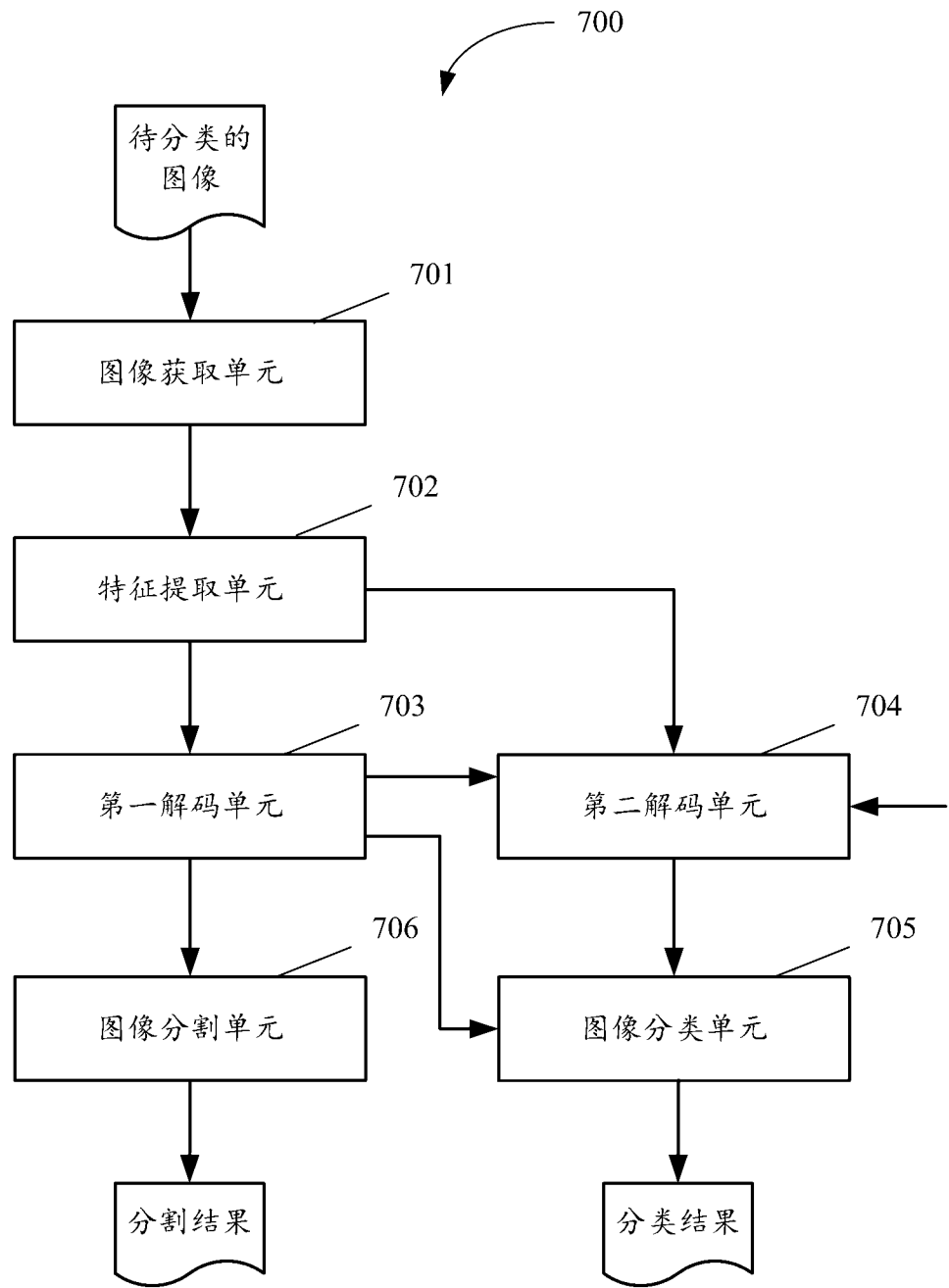


图 7

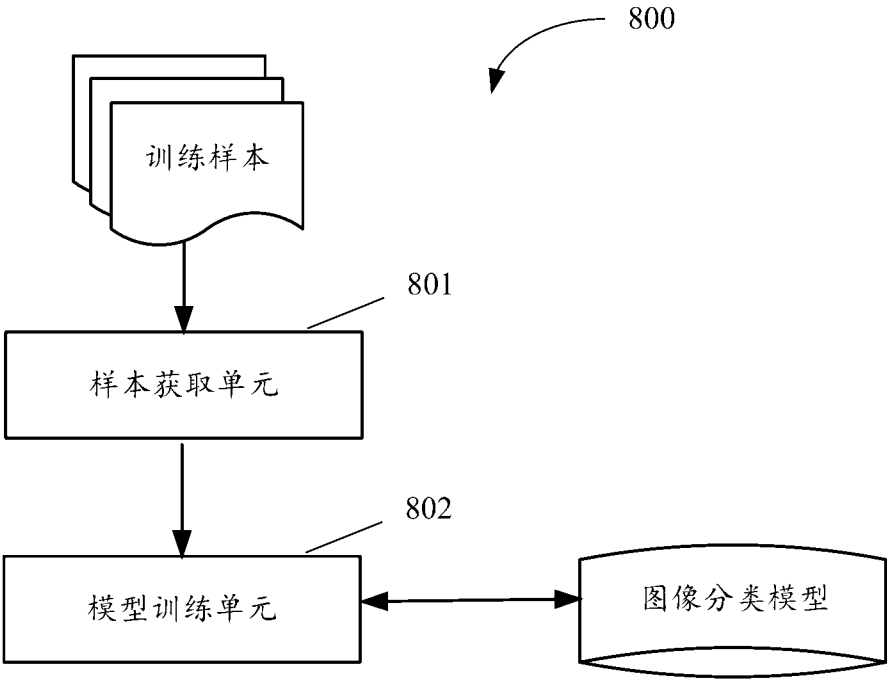


图 8

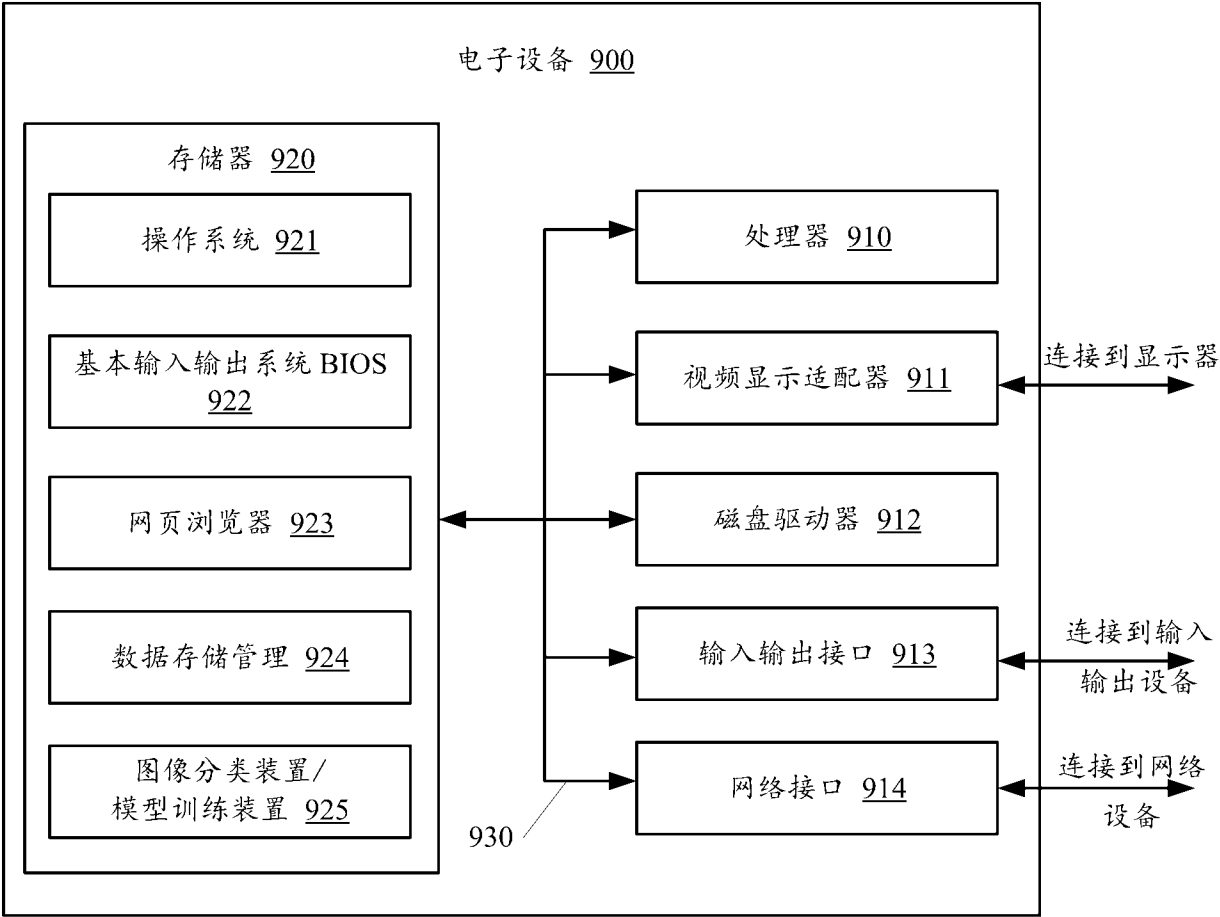


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/103726

A. CLASSIFICATION OF SUBJECT MATTER G06V10/764(2022.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC:G06V,G06F,G06N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 图像, 分类, 特征, 提取, 聚类, 中心, 交叉, 注意力, 编码, 解码, 模型, 矩阵, image, classification, feature, extraction, clustering, center, intersection, attention, encoding, decoding, model, matrix, transformer, token		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 116958659 A (ALIBABA DAMOYUAN (HANGZHOU) TECHNOLOGY CO., LTD.) 27 October 2023 (2023-10-27) claims 1-14, and description, paragraphs [0147]-[0165]	1-24
A	CN 115239675 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 25 October 2022 (2022-10-25) description, paragraphs [0002]-[0051], and abstract	1-24
A	CN 115631183 A (EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY) 20 January 2023 (2023-01-20) entire document	1-24
A	CN 113971735 A (XIDIAN UNIVERSITY) 25 January 2022 (2022-01-25) entire document	1-24
A	US 2023186600 A1 (VINAI ARTIFICIAL INTELLIGENCE APPLICATION AND RESEARCH JOINT STOCK COMPANY) 15 June 2023 (2023-06-15) entire document	1-24
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: “A” document defining the general state of the art which is not considered to be of particular relevance “D” document cited by the applicant in the international application “E” earlier application or patent but published on or after the international filing date “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) “O” document referring to an oral disclosure, use, exhibition or other means “P” document published prior to the international filing date but later than the priority date claimed “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art “&” document member of the same patent family		
Date of the actual completion of the international search 09 September 2024		Date of mailing of the international search report 19 September 2024
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2024/103726

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2022127227 A1 (XI'AN JIAOTONG UNIVERSITY) 23 June 2022 (2022-06-23) entire document	1-24

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2024/103726

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	116958659	A	27 October 2023	None			
CN	115239675	A	25 October 2022	None			
CN	115631183	A	20 January 2023	None			
CN	113971735	A	25 January 2022	None			
US	2023186600	A1	15 June 2023	VN	95351	A	26 June 2023
WO	2022127227	A1	23 June 2022	CN	112529878	A	19 March 2021

A. 主题的分类

G06V10/764(2022.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:G06V,G06F,G06N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

VEN,CNABS,CNTXT,WOTXT,EPTXT,USTXT,CNKI,IEEE:图像, 分类, 特征, 提取, 聚类, 中心, 交叉, 注意力, 编码, 解码, 模型, 矩阵, image, classification, feature, extraction, clustering, center, intersection, attention, encoding, decoding, model, matrix, transformer, token

C. 相关文件

类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 116958659 A (阿里巴巴达摩院(杭州)科技有限公司) 2023年10月27日 (2023 - 10 - 27) 权利要求1-14, 说明书第[0147]-[0165]段	1-24
A	CN 115239675 A (平安科技(深圳)有限公司) 2022年10月25日 (2022 - 10 - 25) 说明书第[0002]-[0051]段, 摘要	1-24
A	CN 115631183 A (华东理工大学) 2023年1月20日 (2023 - 01 - 20) 全文	1-24
A	CN 113971735 A (西安电子科技大学) 2022年1月25日 (2022 - 01 - 25) 全文	1-24
A	US 2023186600 A1 (VINAI ARTIFICIAL INTELLIGENCE APPLICATION AND RESEARCH JOINT STOCK COMPANY) 2023年6月15日 (2023 - 06 - 15) 全文	1-24
A	WO 2022127227 A1 (XI'AN JIAOTONG UNIVERSITY) 2022年6月23日 (2022 - 06 - 23) 全文	1-24

☐ 其余文件在C栏的续页中列出。

☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2024年9月9日

国际检索报告邮寄日期

2024年9月19日

ISA/CN的名称和邮寄地址

中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

吴媛媛

电话号码 (+86) 010-53961351

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2024/103726

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	116958659	A	2023年10月27日	无	
CN	115239675	A	2022年10月25日	无	
CN	115631183	A	2023年1月20日	无	
CN	113971735	A	2022年1月25日	无	
US	2023186600	A1	2023年6月15日	VN	95351 A 2023年6月26日
WO	2022127227	A1	2022年6月23日	CN	112529878 A 2021年3月19日