

直播审核的方法、装置、服务器和存储介质

本申请要求在2020年12月30日提交中国专利局、申请号为202011613190.6的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及互联网资源中违规内容审核领域，例如涉及一种直播审核的方法、装置、服务器和存储介质。

背景技术

随着互联网技术的快速发展，网络上的直播平台也越来越多，使得互联网资源的传播越来越广泛。与此同时，由于一些直播间的直播内容可能会涉及到恐怖、暴力、色情或政治敏感等话题，而使得大量违规的直播内容在互联网上被快速传播出来，因此针对此类直播间，需要实时检测是否存在违规的直播内容，以防止违规直播内容的传播。

通常会利用单特征训练的神经网络模型对直播过程中的多个直播视频帧所体现的直播视觉特征和直播语音下的直播音频特征进行违规内容分析，判断多个直播视频帧中是否存在违规画面，或者判断直播语音中是否存在涉及恐怖、暴力、色情或政治敏感等话题的违规音频，从而审核出可疑的违规直播间，推送给人工审核平台进行人工审核；此时，在直播过程中提取直播视频帧和直播语音时，会存在由于直播网络不稳定而导致提取结果不准确的问题，因此在利用所提取的直播视频帧和直播语音来审核直播间内是否存在违规内容时，无法保证直播间违规审核的准确率和召回率，极易造成误判或漏判的问题。

发明内容

本申请提供了一种直播审核的方法、装置、服务器和存储介质，避免直播审核下存在的误审核和漏审核，提高直播审核的准确率和召回率。

提供了一种直播审核的方法，该方法包括：

通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核；

如果所述当前直播帧通过初步违规审核，则将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到预先构建的行为审核模型中，

得到所述当前直播帧的目标违规得分。

还提供了一种直播审核的装置，该装置包括：

初步审核模块，设置为通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核；

违规得分确定模块，设置为如果所述当前直播帧通过初步违规审核，则将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到预先构建的行为审核模型中，得到所述当前直播帧的目标违规得分。

还提供了一种服务器，该服务器包括：

一个或多个处理器；

存储装置，设置为存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现本申请任意实施例所述的直播审核的方法。

还提供了一种计算机可读存储介质，存储有计算机程序，该计算机程序被处理器执行时实现本申请任意实施例所述的直播审核的方法。

附图说明

图 1A 为本申请实施例一提供的一种直播审核的方法的流程图；

图 1B 为本申请实施例一提供的直播审核过程的原理示意图；

图 2A 为本申请实施例二提供的一种直播审核的方法的流程图；

图 2B 为本申请实施例二提供的直播审核过程的原理示意图；

图 3A 为本申请实施例三提供的一种直播审核的方法的流程图；

图 3B 为本申请实施例三提供的直播审核过程的原理示意图；

图 4 为本申请实施例四提供的一种直播审核的装置的结构示意图；

图 5 为本申请实施例五提供的一种服务器的结构示意图。

具体实施方式

下面结合附图和实施例对本申请进行说明。此处所描述的实施例仅仅用于

解释本申请，而非对本申请的限定。为了便于描述，附图中仅示出了与本申请相关的部分而非全部结构。此外，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

实施例一

图 1A 为本申请实施例一提供的一种直播审核的方法的流程图，本实施例可适用于对任一个直播间内的直播画面进行违规审核的情况中。本实施例提供了一种直播审核的方法可以由本申请实施例提供的直播审核的装置来执行，该装置可以通过软件和/或硬件的方式来实现，并集成在执行本方法的服务器中，该服务器可以是承载有视频审核能力的后台服务端。

参考图 1A，该方法可以包括如下步骤。

S110，通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核。

由于直播间面向的观众类型较为广泛，如果一些直播间的直播内容涉及到恐怖、暴力、色情或政治敏感等话题，那么违规内容就会在互联网中被快速传播起来，因此为了防止违规内容对用户生活产生的不利影响，需要在直播过程中，实时检测是否存在违规的直播内容。

在本实施例中，为了保证直播违规审核的全面性，会将直播审核下的整个审核过程分成初步违规审核和再次违规审核两部分。此时，在初步违规审核过程中，能够快速审核出容易区分是否违规的较为明显的违规直播，而在再次违规审核过程中，采用相比初步违规审核来说更高的审核力度，来对初步违规审核无法准确判断是否存在违规内容的直播帧再次进行违规审核，以避免直播审核下存在误审核和漏审核的问题，从而保证待审核直播间内当前直播帧的全面违规审核。

本实施例在初步违规审核过程中，会预先构建出高精确审核模型和高召回审核模型两个网络模型；其中，高精确审核模型具备高准确性的违规审核能力，在通过高精确审核模型对待审核直播间内的当前直播帧进行违规审核时，能够保证所审核出的违规直播帧极大可能是真实违规的，避免出现违规误审核的问题；同时，高召回审核模型具备高召回性的违规审核能力，在通过高召回审核模型对待审核直播间内的当前直播帧进行违规审核时，能够尽可能保证即使当前直播帧存在极小的违规内容，也能够被审核出来，减少出现违规漏审核的情况。

本实施例在对高精确审核模型和高召回审核模型进行训练时，会预先构建一个训练样本集，该训练样本集中会包括多类直播间内历史直播视频中的历史

直播帧，同时按照每个历史直播帧是否真实违规，以及被准确审核出是否违规的难易程度，可以将训练样本集内的每个训练样本分为易正样本、易负样本、难正样本和难负样本四类，易正样本和易负样本分别是指容易被模型识别出是否违规的正样本和负样本，难正样本和难负样本分别是指难以被模型识别出是否违规的正样本和负样本，此时由于高精确审核模型对于违规审核的精确性具有较高要求，因此该高精确审核模型的训练样本可以由训练样本集中的易正样本和易负样本构成，通过审核易正样本和易负样本是否存在违规内容，不断优化该高精确审核模型的模型参数和网络结构等；而由于高召回审核模型对于违规审核的召回性具有较高要求，因此该高召回审核模型的训练样本可以由训练样本集中的难正样本和难负样本构成，通过采用难以被审核是否违规的训练样本进行审核训练，使得该高召回审核模型面对难以被审核出来的违规直播帧，也能够审核出来，从而保证高召回审核模型对于违规审核的高召回能力。

同时，由于高精确审核模型只针对违规审核精确性进行训练，存在较大的违规漏审核的情况，而高召回审核模型却具备高召回的违规审核能力，能够尽可能的减少违规漏审核的情况，因此为了保证初步违规审核的高准确性和高召回性，本实施例可以将高精确审核模型和高召回审核模型级联起来，如图 1B 所示，从而先利用高精确审核模型对待审核直播间内的每一个当前直播帧进行初始的违规审核，而对高精确审核模型未审核出是否违规的当前直播帧，则由高召回审核模型继续进行违规审核，以采用级联的两层审核模型依次进行违规审核，完成本实施例中的初步违规审核过程，保证初步违规审核的高准确性和高召回性。

在待审核直播间的视频直播过程中，本实施例可以每隔预设时长（如每隔 2s 等）从待审核直播间内截取一张直播图片，作为本实施例中的当前直播帧，将该当前直播帧输入到两层级联网络中的高精确审核模型中，由该高精确审核模型判断该当前直播帧内是否存在违规内容，若该当前直播帧通过该高精确审核模型的违规审核，说明该高精确审核模型认为该当前直播帧内不存在违规内容，而由于高精确审核模型存在违规漏审核的情况，因此需要再次将该当前直播帧输入到两层级联网络中的高召回审核模型中，由该高召回审核模型继续判断该当前直播帧内是否存在违规内容，从而完成对待审核直播间内当前直播帧的初步违规审核。

此时，高召回审核模型对于该当前直播帧的违规审核结果存在两种情况：1）未通过高召回审核模型的违规审核，说明该高召回审核模型能够确定该当前直播帧中存在违规内容，直接确认该当前直播帧所在的待审核直播间为违规直播间；2）通过高召回审核模型的违规审核，说明该高召回审核模型认为该当前直播帧内不存在违规内容，则后续需要利用预先构建的行为审核模型继续对该当

前直播帧进行违规审核。

S120, 如果当前直播帧通过初步违规审核, 则将当前直播帧在高精确审核模型下的第一违规得分、在高召回审核模型下的第二违规得分以及在待审核直播间内的多维行为特征, 输入到预先构建的行为审核模型中, 得到当前直播帧的目标违规得分。

可选的, 如果待审核直播间内的当前直播帧既通过高精确审核模型的违规审核, 也通过高召回审核模型的违规审核, 则说明当前直播帧通过初步违规审核, 需要进行再次违规审核, 以保证违规审核的准确性和全面性。

通过级联的高精确审核模型和高召回审核模型, 对当前直播帧进行初步违规审核时, 通常是对当前直播帧的直播视觉画面内的信息进行违规分析, 此时高精确审核模型会输出当前直播帧的第一违规得分, 而高召回审核模型会输出当前直播帧的第二违规得分, 且该第一违规得分和第二违规得分均能够代表当前直播帧的视觉画面特征。

然而, 由于直播间内的直播内容具有多变性, 且直播间内的主播与观众之间也会针对直播内容进行互动, 例如直播打赏、内容公屏评论和主播历史违规情况等, 使得直播间内的互动行为也能够作为判断当前直播帧是否存在违规内容的参考依据, 因此本实施例在确定当前直播帧通过初步违规审核后, 在对当前直播帧进行再次违规审核时, 除了可以参考当前直播帧在高精确审核模型下的第一违规得分和在高召回审核模型下的第二违规得分所代表的视觉画面特征外, 还会通过分析当前直播帧在待审核直播间内所处的直播时刻下, 主播与观众在该待审核直播间内的多项直播互动行为, 确定当前直播帧在该待审核直播间内的多维行为特征, 如图 1B 所示, 并将当前直播帧的第一违规得分和第二违规得分所代表的视觉画面特征和由待审核直播间内的多项直播互动行为所确定的多维行为特征结合起来, 共同输入到预先构建的行为审核模型中, 该行为审核模型能够对第一违规得分和第二违规得分所代表的视觉画面特征和由待审核直播间内的多项直播互动行为所确定的多维行为特征进行融合分析, 统一判断该当前直播帧内是否存在违规内容, 从而输出当前直播帧的目标违规得分, 此时该目标违规得分采用由视觉画面特征的多维行为特征相结合的多角度特征融合分析, 来实现再次违规审核, 无需采用额外的审核设备, 降低了违规审核成本, 保证直播间违规审核的准确性和召回性。

本实施例提供的技术方案, 通过级联的高精确审核模型和高召回审核模型, 对待审核直播间内的当前直播帧进行初步违规审核, 能够快速审核出容易区分和违规明显的违规直播帧, 而对于通过初步违规审核的不易区分或者违规不明显的当前直播帧, 会在该当前直播帧在高精确审核模型下的第一违规得分和在

高召回审核模型下的第二违规得分的基础上，结合该当前直播帧在待审核直播间内的多维行为特征，通过行为审核模型再次对当前直播帧进行视觉画面特征和行为特征相结合的统一特征分析，从而得到该当前直播帧的目标违规得分，实现待审核直播间内当前直播帧的全面违规审核，避免直播审核下存在误审核和漏审核的问题，在降低违规审核成本的基础上，提高直播审核的准确率和召回率。

实施例二

图 2A 为本申请实施例二提供的一种直播审核的方法的流程图，图 2B 为本申请实施例二提供的直播审核过程的原理示意图。本实施例是在上述实施例的基础上进行说明。如图 2A 所示，本实施例中对于待审核直播间内当前直播帧的再次违规审核的审核过程和行为审核模型的建模过程进行解释说明。

可选的，如图 2A 所示，本实施例中可以包括如下步骤。

S210，通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核。

S220，如果当前直播帧通过初步违规审核，则合并当前直播帧在高精确审核模型下的第一违规得分、在高召回审核模型下的第二违规得分以及在待审核直播间内的多维行为特征，得到当前直播帧的直播行为特征。

可选的，如果当前直播帧通过初步违规审核，说明高精确审核模型和高召回审核模型都认为当前直播帧内不存在违规内容，因此为了保证直播审核的准确性，避免出现违规漏审核的情况，还会利用当前直播帧的视觉画面特征和行为特征共同对该当前直播帧继续进行再次违规审核，此时会确定出当前直播帧在高精确审核模型下的第一违规得分和高召回审核模型下的第二违规得分，将第一违规得分和第二违规得分作为当前直播帧对应的视觉画面特征，同时根据当前直播帧的直播时刻下，在待审核直播间内的多项直播互动行为，确定出当前直播帧在待审核直播间内的多维行为特征，将第一违规得分、第二违规得分和多维行为特征合并起来，得到当前直播帧的直播行为特征，作为再次违规审核下的多角度融合特征，后续采用该直播行为特征判断当前直播帧内是否存在违规内容，能够保证直播审核的全面性和准确性。

示例性的，本实施例中的直播行为特征可以为 $X=(x1, x2, \dots, x11)$ ，其中 $x1$ 为当前直播帧在高精确审核模型下的第一违规得分， $x2$ 为当前直播帧在高召回审核模型下的第二违规得分， $x3$ 为多维行为特征中在待审核直播间内的直播实时观看数目， $x4$ 为多维行为特征中在待审核直播间内的用户发送公屏数量，

x5 为多维行为特征中在待审核直播间内的违规公屏数量, x6 为多维行为特征中在待审核直播间内违规公屏在所有公屏中的占比, x7 为多维行为特征中在待审核直播间内的直播打赏数量, x8 为多维行为特征中在待审核直播间内的主播历史违规次数, x9 为多维行为特征中在待审核直播间内的主播历史违规占比, x10 为多维行为特征中在待审核直播间内的主播历史语音违规次数, x11 为多维行为特征中在待审核直播间内的主播历史语音违规占比。

为了保证直播违规审核的准确性, 本实施例中的多维行为特征可以动态配置, 对此不作限定。

S230, 将直播行为特征输入到行为审核模型中, 得到当前直播帧的目标违规得分。

可选的, 在得到直播行为特征后, 可以将该直播行为特征作为多角度下的违规融合特征, 输入到预先构建的行为审核模型中, 通过该行为审核模型对该直播行为特征在每一个维度下的特征值进行融合分析, 判断是否携带违规内容信息, 进而输出当前直播帧的目标违规得分, 通过该目标违规得分判断当前直播帧内是否存在违规内容, 提高直播审核的准确性。

示例性的, 本实施例可以采用如下步骤对行为审核模型进行建模: 从历史直播视频集内的每一个历史直播视频中分别提取对应的历史直播帧, 并确定每一个历史直播帧在所处直播间内的多维历史行为特征以及该历史直播帧的违规标签; 针对每一个历史直播帧, 将该历史直播帧在高精确审核模型下的第一历史违规得分、在高召回审核模型下的第二历史违规得分以及在所处直播间内的多维历史行为特征, 输入到预先构建的行为审核模型中, 得到历史直播帧的历史违规得分; 以最小化每一个历史直播帧的历史违规得分和违规标签之间的差异为目标, 不断优化行为审核模型。

通过查找多个直播间的历史直播视频, 构建出用于建模的历史直播视频集, 而由于对直播间进行违规审核时, 主要是对直播过程中的直播帧进行违规审核, 因此需要对历史直播视频集内的每一个历史直播视频进行视频帧提取, 例如从每一个历史直播视频内每隔预设时长便截取一张历史直播图片, 作为从该历史直播视频内所提取出的历史直播帧, 此时所提取出的全部历史直播帧可以构成该行为审核模型的建模样本。

此时, 为了保证行为审核模型的成功建模, 需要设定建模过程中所需要参考的多个建模特征, 以及该行为审核模型的建模优化目标, 以按照该建模优化目标对多个建模特征下的实际建模情况进行不断优化, 例如本实施例中的建模特征可以为 $X=(x_1, x_2, \dots, x_{11})$, 其中 x_1 为历史直播帧在高精确审核模型下的第一历史违规得分, x_2 为历史直播帧在高召回审核模型下的第二历史违规得

分, x_3 为所设定的多维历史行为特征中历史直播帧在所在直播间内的直播实时观看数目, x_4 为多维历史行为特征中历史直播帧在所在直播间内的用户发送公屏数量, x_5 为多维历史行为特征中历史直播帧在所在直播间内的违规公屏数量, x_6 为多维历史行为特征中历史直播帧在所在直播间内违规公屏在所有公屏中的占比, x_7 为多维历史行为特征中历史直播帧在所在直播间内的直播打赏数量, x_8 为多维历史行为特征中历史直播帧在所在直播间内的主播历史违规次数, x_9 为多维历史行为特征中历史直播帧在所在直播间内的主播历史违规占比, x_{10} 为多维历史行为特征中历史直播帧在所在直播间内的主播历史语音违规次数, x_{11} 为多维历史行为特征中历史直播帧在所在直播间内的主播历史语音违规占比。

由于多个建模特征中得历史直播帧在所处直播间内的多维历史行为特征在历史直播过程中就可以确定, 因此在从每一个历史直播视频中分别提取出对应的历史直播帧后, 通过判断历史直播帧在所处直播间内的历史直播互动行为, 即可确定出每一个历史直播帧在所处直播间内的多维历史行为特征, 同时还能够确定每一个历史直播帧的违规标签, 如图 2B 所示, 以便后续按照建模优化目标进行不断建模优化。针对每一个历史直播帧, 可以将该历史直播帧分别输入到本实施例的高精度审核模型和高召回审核模型中, 从而得到每一个历史直播帧的第一历史违规得分和第二历史违规得分, 以此即可将每一个历史直播帧的第一历史违规得分、第二历史违规得分和在所处直播间内的多维历史行为特征, 共同组成每一个历史直播帧对应的建模特征, 并将每一个历史直播帧对应的建模特征不断输入到该行为审核模型中, 得到每一个历史直播帧的历史违规得分。

此时, 本实施例中行为审核模型的建模优化目标可以为以最小化每一个历史直播帧的历史违规得分和违规标签之间的差异为目标, 因此在建模优化过程中不断分析每一个历史直播帧的历史违规得分与每一个历史直播帧的违规标签之间的差异, 并通过优化行为审核模型中的数学参数来不断减小该差异, 以达到历史直播帧的历史违规得分和违规标签之间的差异能够最小化的目标, 从而得到最终优化后的行为审核模型, 此时最终优化后的行为审核模型具备高准确性的违规审核能力。

示例性的, 本实施例中行为审核模型的建模优化目标可以为:

$$\min \sum_{x^i \in V} \left(\hat{y}^i - y^i \right)^2 \quad s.t. \hat{y}^i = p(x^i) ; \text{ 其中, } x^i \text{ 为第 } i \text{ 个历史直播帧的建模特征, } y^i \text{ 为第 } i \text{ 个历史直播帧的违规标签, } y^i = 1 \text{ 表示存在违规内容, } y^i = 0 \text{ 表示不存在违规内容, } \hat{y}^i \text{ 为通过行为审核模型对第 } i \text{ 个历史直播帧确定的历史违规得分,}$$

$V = (x^1, x^2, \dots, x^m)$, 表示历史直播帧组成的建模样本, m 为样本数量, $p(x^i)$ 表示 x^i 通过行为审核模型得到的预测概率。

本实施例提供的技术方案, 以最小化每一个历史直播帧的历史违规得分和违规标签之间的差异为建模优化目标, 对行为审核模型进行建模, 保证行为审核模型进行违规审核的准确性, 后续通过合并待审核直播间内当前直播帧在高精确审核模型下的第一违规得分、在高召回审核模型下的第二违规得分以及在待审核直播间内的多维行为特征, 得到当前直播帧的直播行为特征, 并通过建模后的行为审核模型对该直播行为特征进行多角度下的违规特征分析, 提高直播审核的准确率和召回率。

实施例三

图 3A 为本申请实施例三提供的一种直播审核的方法的流程图, 图 3B 为本申请实施例三提供的直播审核过程的原理示意图。本实施例是在上述实施例的基础上进行说明。如图 3A 所示, 本实施例中对待审核直播间的整体违规审核过程进行解释说明。

可选的, 如图 3A 所示, 本实施例中可以包括如下步骤。

S310, 将待审核直播间内的当前直播帧输入到预先构建的高精确审核模型中, 得到当前直播帧的第一违规得分。

可选的, 在对待审核直播间进行违规审核时, 会在直播过程中实时检测直播画面是否存在违规内容, 因此需要实时对直播过程中的当前直播帧进行违规审核。将当前直播帧输入到预先构建的高精确审核模型中, 通过该高精确审核模型对当前直播帧内的直播画面进行初始的违规审核, 从而输出该当前直播帧的第一违规得分, 后续通过判断该第一违规得分是否超出预设精确违规阈值, 能够确定该当前直播帧内是否存在违规内容。

S320, 判断第一违规得分是否超出预设精确违规阈值, 若是, 执行 S380; 若否, 执行 S330。

S330, 将当前直播帧继续输入到预先构建的高召回审核模型中, 得到当前直播帧的第二违规得分。

可选的, 如果第一违规得分未超出预设精确违规阈值, 说明高精确审核模型认为该当前直播帧内不存在违规内容, 然而由于高精确审核模型存在违规漏审核的问题, 因此需要将当前直播帧继续输入到预先构建的高召回审核模型中,

如图 3B 所示,通过该高召回审核模型继续对当前直播帧内的直播画面进行相应的违规审核,从而输出该当前直播帧的第二违规得分,后续通过判断该第二违规得分是否超出预设召回违规阈值,能够再次确定该当前直播帧内是否存在违规内容。

S340,判断第二违规得分是否超出预设召回违规阈值,若是,执行 S380;若否,执行 S350。

如果第二违规得分未超出预设召回违规阈值,说明高召回审核模型也认为该当前直播帧内不存在违规内容,也就是当前直播帧同时通过高精确审核模型和高召回审核模型的违规审核,也就是通过初步违规审核,利用行为审核模型继续对当前直播帧进行再次违规审核。所述当前直播帧通过所述初步违规审核为所述当前直播帧的第一违规得分未超出所述预设精确违规阈值,且所述当前直播帧的第二违规得分未超出预设召回违规阈值。

S350,将当前直播帧在所述高精确审核模型下的第一违规得分、在高召回审核模型下的第二违规得分以及在待审核直播间内的多维行为特征,输入到预先构建的行为审核模型中,得到当前直播帧的目标违规得分。

S360,判断目标违规得分是否超出预设行为违规阈值,若是,执行 S380;若否,执行 S370。

可选的,在利用行为审核模型输出当前直播帧的目标违规得分后,通过判断目标违规得分是否超出预设行为违规阈值,可以最终审核出当前直播帧内是否存在违规内容,以便执行不同的直播处理流程。

S370,间隔预设直播时长后,从待审核直播间内采集新的当前直播帧,继续对新的当前直播帧进行违规审核。

可选的,如果目标违规得分未超出预设行为违规阈值,说明行为审核模型也认为当前直播帧内不存在违规内容,此时本实施例中的三个模型均认为当前直播帧内不存在违规内容,则可以确定当前直播帧为非违规帧,说明待审核直播间内当前不存在违规内容,此时为了保证直播间违规审核的实时性,会在间隔预设直播时长(如每隔 2s)后,继续从待审核直播间内的当前直播时刻下截取一张新的直播图片,作为所采集的新的当前直播帧,按照本实施例 S310-S380 的步骤继续对该新的当前直播帧进行违规审核,从而对待审核直播间进行实时违规审核,直至直播结束或者审核出违规的直播帧。

S380,将待审核直播间的直播信息推送给人工审核平台。

可选的,在第一违规得分超出预设精确违规阈值,或者第二违规得分超出预设召回违规阈值,或者目标违规得分超出预设行为违规阈值时,都说明当前

审核使用的模型认为该当前直播帧内存在违规内容，但是为了避免出现违规误审核的情况，如图 3B 所示，本实施例还会额外设置一个人工审核平台，以在审核出当前直播帧内存在违规内容时，将该当前直播帧所在的待审核直播间的直播信息推送给人工审核平台，以进一步对该待审核直播间的直播内容进行人工审核，从而保证直播审核的准确性。

本实施例中待审核直播间的直播信息可以为待审核直播间的直播地址或正在直播的视频画面等，对此不作限定。

S390，实时监控待审核直播间内的直播互动项，如果直播互动项超出预设互动阈值，则将待审核直播间的直播信息推送给人工审核平台，并停止对待审核直播间内当前直播帧的违规审核。

可选的，为了避免出现违规审核的异常情况，而保证直播审核的高召回率，本实施例在通过构建模型来实现直播违规审核的基础上，还会通过实时监控待审核直播间内的直播互动项，来判断是否需要直接推送给人工审核平台进行人工审核，将人工审核作为直播审核的审核异常保底方案。此时，该直播互动项可以包括待审核直播间内的直播观看数量和直播打赏数量等，以此分析待审核直播间的受众广泛程度，如果直播互动项超出预设互动阈值，说明待审核直播间的直播受众率很大，一旦出现违规内容，则会快速传播出去，因此需要直接将待审核直播间的直播信息推送给人工审核平台进行人工审核，以确保直播违规审核的高效性，同时停止利用所构建的多个模型对待审核直播间内当前直播帧的违规审核。

本实施例提供的技术方案，在确定当前直播帧内存在违规内容后，直接将待审核直播间的直播信息推送给人工审核平台进行人工审核，避免出现违规误审核的情况，提高直播违规审核的准确性；同时，利用待审核直播间内的直播互动项额外设置直播审核的异常保底方案，保证直播违规审核的全面性。

实施例四

图 4 为本申请实施例四提供的一种直播审核的装置的结构示意图。如图 4 所示，该装置可以包括：

初步审核模块 410，设置为通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核；

违规得分确定模块 420，设置为如果所述当前直播帧通过初步违规审核，则将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到

预先构建的行为审核模型中，得到所述当前直播帧的目标违规得分。

本实施例提供的技术方案，通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核，能够快速审核出容易区分和违规明显的违规直播帧，而对于通过初步违规审核的不易区分或者违规不明显的当前直播帧，会在该当前直播帧在高精确审核模型下的第一违规得分和在高召回审核模型下的第二违规得分的基础上，结合该当前直播帧在待审核直播间内的多维行为特征，通过行为审核模型再次对当前直播帧进行视觉画面特征和行为特征相结合的统一特征分析，从而得到该当前直播帧的目标违规得分，实现待审核直播间内当前直播帧的全面违规审核，避免直播审核下存在误审核和漏审核的问题，在降低违规审核成本的基础上，提高直播审核的准确率和召回率。

本实施例提供的直播审核的装置可适用于执行上述任意实施例提供的直播审核的方法，具备相应的功能和效果。

实施例五

图 5 为本申请实施例五提供的一种服务器的结构示意图，如图 5 所示，该服务器包括处理器 50、存储装置 51 和通信装置 52；服务器中处理器 50 的数量可以是一个或多个，图 5 中以一个处理器 50 为例；服务器中的处理器 50、存储装置 51 和通信装置 52 可以通过总线或其他方式连接，图 5 中以通过总线连接为例。

本实施例提供的一种服务器可用于执行上述任意实施例提供的直播审核的方法，具备相应的功能和效果。

实施例六

本申请实施例六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，该计算机程序被处理器执行时可实现上述任意实施例中的直播审核的方法。该方法可以包括：

通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核；

如果所述当前直播帧通过初步违规审核，则将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到预先构建的行为审核模型中，

得到所述当前直播帧的目标违规得分。

本申请实施例所提供的一种包含计算机可执行指令的存储介质，其计算机可执行指令不限于如上所述的方法操作，还可以执行本申请任意实施例所提供的直播审核的方法中的相关操作。存储介质可以是非暂态（non-transitory）存储介质。

本申请可借助软件及通用硬件来实现，也可以通过硬件实现。本申请的技术方案可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如计算机的软盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、闪存（FLASH）、硬盘或光盘等，包括多个指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等等）执行本申请多个实施例所述的方法。

上述直播审核的装置的实施例中，所包括的多个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，多个功能单元的名称也只是为了便于相互区分，并不用于限制本申请的保护范围。

权利要求书

1、一种直播审核的方法，包括：

通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核；

在所述当前直播帧通过所述初步违规审核的情况下，将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到预先构建的行为审核模型中，得到所述当前直播帧的目标违规得分。

2、根据权利要求1所述的方法，其中，所述将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到预先构建的行为审核模型中，得到所述当前直播帧的目标违规得分，包括：

合并所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，得到所述当前直播帧的直播行为特征；

将所述直播行为特征输入到所述行为审核模型中，得到所述当前直播帧的目标违规得分。

3、根据权利要求1所述的方法，其中，所述行为审核模型通过执行如下步骤建模：

从历史直播视频集内的每一个历史直播视频中提取对应的历史直播帧，并确定每一个历史直播帧在所处直播间内的多维历史行为特征以及所述历史直播帧的违规标签；

针对每一个历史直播帧，将所述历史直播帧在所述高精确审核模型下的第一历史违规得分、在所述高召回审核模型下的第二历史违规得分以及在所处直播间内的多维历史行为特征，输入到预先构建的行为审核模型中，得到所述历史直播帧的历史违规得分；

以最小化每一个历史直播帧的历史违规得分和违规标签之间的差异为目标，不断优化所述行为审核模型。

4、根据权利要求1所述的方法，其中，在所述将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到预先构建的行为审核模型中，得到所述当前直播帧的目标违规得分之后，还包括：

在所述目标违规得分超出预设行为违规阈值的情况下，将所述待审核直播

间的直播信息推送给人工审核平台；

在所述目标违规得分未超出所述预设行为违规阈值的情况下，间隔预设直播时长后，从所述待审核直播间内采集新的当前直播帧，继续对所述新的当前直播帧进行违规审核。

5、根据权利要求4所述的方法，还包括：

实时监控所述待审核直播间内的直播互动项，在所述直播互动项超出预设互动阈值的情况下，将所述待审核直播间的直播信息推送给所述人工审核平台，并停止对所述待审核直播间内当前直播帧的违规审核。

6、根据权利要求1所述的方法，其中，所述通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核，包括：

将所述待审核直播间内的当前直播帧输入到预先构建的高精确审核模型中，得到所述当前直播帧的第一违规得分；

在所述第一违规得分未超出预设精确违规阈值的情况下，将所述当前直播帧继续输入到预先构建的高召回审核模型中，得到所述当前直播帧的第二违规得分。

7、根据权利要求6所述的方法，其中，所述当前直播帧通过所述初步违规审核为所述当前直播帧的第一违规得分未超出所述预设精确违规阈值，且所述当前直播帧的第二违规得分未超出预设召回违规阈值。

8、根据权利要求6所述的方法，还包括：

在所述第一违规得分超出所述预设精确违规阈值，或者所述第二违规得分超出预设召回违规阈值的情况下，将所述待审核直播间的直播信息推送给人工审核平台。

9、根据权利要求1-8中任一项所述的方法，其中，所述高精确审核模型的训练样本由训练样本集中的易正样本和易负样本构成，所述高召回审核模型的训练样本由训练样本集中的难正样本和难负样本构成。

10、一种直播审核的装置，包括：

初步审核模块，设置为通过级联的高精确审核模型和高召回审核模型，对待审核直播间内的当前直播帧进行初步违规审核；

违规得分确定模块，设置为在所述当前直播帧通过所述初步违规审核的情况下，将所述当前直播帧在所述高精确审核模型下的第一违规得分、在所述高召回审核模型下的第二违规得分以及在所述待审核直播间内的多维行为特征，输入到预先构建的行为审核模型中，得到所述当前直播帧的目标违规得分。

11、一种服务器，包括：

一个或多个处理器；

存储装置，设置为存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-9 中任一项所述的直播审核的方法。

12、一种计算机可读存储介质，存储有计算机程序，其中，所述计算机程序被处理器执行时实现如权利要求 1-9 中任一项所述的直播审核的方法。