阿里云
内容安全

用户指南

文档版本：20180911

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的”现状“、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含”阿里云”、Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **禁止：** 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **警告：** 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 **说明：** 您也可以通过按**Ctrl** + **A**选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 **确定**。
`courier` 字体	命令。	执行 `cd /d C:/windows` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig` *`[-all|-t]`*
{}或者{a|b}	表示必选项，至多选择一个。	`swich` *`{stand | slave}`*

目录

1 使用限制

本文描述了您在使用内容安全时应该注意的产品功能限制。

站点检测

站点检测的对象是您的网站上的网页和图片，以URL数量进行计数。在单个网站的一个检测周期内，站点检测支持的最大检测容量为10万个URL。

OSS违规检测

- OSS违规检测只向阿里云OSS用户提供服务。使用该服务前，您需要在内容安全控制台，通过RAM，授权内容服务读取OSS Bucket的权限（该操作在您首次登录控制台时，可以一键完成）。
- 增量图片支持自动检测，但需要您在内容安全设置中添加具体的 Bucket；这样，内容安全才会每日帮助您检测该 Bucket 内实时增加的图片。
- 存量图片不支持自动检测，如果您想要对存量图片进行检测，您需要手动选择 Bucket 和时间范围，进行存量扫描。

2 站点检测

站点检测服务定期检查您的网站首页和全站内容，及时发现您的网站在内容安全方面可能存在的风险（如首页篡改，挂马暗链，色情低俗，涉政暴恐等），并向您展示违规内容的具体地址，帮助您查看和修复。您可以设置消息通知，选择邮件、短信、站内信的方式，获取实时的站点首页风险提醒。

功能描述

购买站点检测实例后，您需要将实例绑定到您的站点，添加要检测的网站域名和首页地址，设定首页和全站检测的频率，并完成网站鉴权。完成设置后，系统将定期按照您设定的频率对首页和全站内容（包含网页源码，文本和图片）进行检测。如果发现有风险，会按照您设定的消息接收方式通知您，您也可以登录产品控制台查看检测结果。

说明：

在对一个站点进行全站检测时，一个站点检测实例在一个检测周期内，支持的最大检测限制为：网页数和图片数合计不超过10万条/张。

站点检测提供以下功能：

- **首页监测**

  定期对您网站的首页进行监测，展示最近一次的检查结果，涵盖首页篡改，挂马暗链，色情低俗，涉政暴恐风险；提供源码，文本，图片三类呈现方式，供您参照和整改。

- **全站检测**

  定期对您网站域名下的网页进行自动化全站内容检测，展示最近一次的检查结果，涵盖挂马暗链，色情低俗，涉政暴恐风险；提供源码，文本，图片三类呈现方式，供您参照和整改。

站点检测支持以下设置：

- **设置首页防篡改基准**

  通过算法对比网页实时状态和您预设基准状态对比，判断是否为非法篡改。

- **添加重点监控URL**

  为了确保全站检测时重要页面不会遗漏，建议您添加网站重点监控URL。最多支持添加5,000条重点监控URL。

- **自定义词库和图库**

在使用站点检测服务时，您可以添加自定义关键词进行黑名单防控；添加的关键词会在15分钟内生效，关键词只支持UTF-8格式。在使用站点检测进行鉴黄、暴恐等图像服务时，您可以添加自定义图片进行黑名单/白名单防控；添加的图片会在15分钟内生效。

关于该功能的更多介绍，请参考文本反垃圾*API*。

前提条件

已购买站点检测实例。

> **说明：**
> 购买方法请参考购买站点检测实例。

操作流程

在购买站点检测实例后，您需要将实例绑定到待检测的站点，为站点启用检测服务。然后，您可以在控制台查看首页检测和全站检测的结果。

如果您想进一步保障检测效果，您可以设置首页防篡改基准和重点监控*URL*，或者自定义词库，自定义图库。

您也可以使用消息通知设置，设置风险通知方式，开启/关闭首页风险实时通知。

启用站点检测

参照以下步骤，为您的站点启用站点检测服务：

1. 登录云盾内容安全控制台。
2. 前往**设置** > **站点检测**页面。
3. 选择一个有效的，处于**未绑定**状态的实例，单击其操作列下的**绑定站点**。

4. 在**绑定站点**对话框中，填写网站信息和检测频率。配置说明如下：

配置	说明
站点协议	勾选**HTTP**或者**HTTPS**。如果您的网站HTTP和HTTPS分别对应不同的内容，而且内容差异较大，建议您绑定2个不同的实例。
站点域名	填写您站点的域名，填写时不要包含`http://`或者`https://`。如果您的网站有多个频道子域名，建议您在这里填写根域名。 **说明：** 假设您的网站有www.domain.com，news.domain.com，sports.domain.com等多个频道内容需要检测，建议您在这里输入domain.com。如果您只想检测news频道的内容news.domain.com，您可以输入news.domain.com。
站点首页地址	填写完整的站点首页网址。输入的网址必须在您要绑定的域名下。
首页监测间隔	设置每隔多少小时，访问您的网站首页进行一次检测。默认为1个小时。
全站检测频率	设置执行全站检测的频率，**7天1次**或者**1天1次**。默认为7天1次。 **说明：** 站点检测频率越高，检测占用的带宽及产生的带宽费用也越多。如果您的网站内容比较多，且网站带宽不足的话，过高的检测频率可能影响您网站的正常访问速度。如果您不希望影响网站性能，建议您配置较低的检测频率。

5. 单击下一步，进行站点验证，证明您对站点的管理权，防止未经授权的检测。
6. 在验证站点对话框，选择验证方式，参照对话框中的验证说明完成相应操作，然后单击立即验证。如果您暂时不方便进行验证，您可以单击**稍后验证**，保存当前已输入的数据。支持的验证方式包括以下四种：

- **阿里云账户验证**：验证待检测站点（域名）是否在您当前登录的阿里云账号的资产下。
- **主机文件验证**：按要求在域名对应主机的根目录下生成相应的文件进行验证。
- **CNAME验证**：按要求在待检测域名的解析记录中增加指定的CNAME记录进行验证。关于添加CNAME记录的操作，请参考添加*CNAME*记录。
- **网站首页HTML标签验证**：按要求修改网站首页HTML源文件进行验证。

7. 验证通过后，完成站点绑定和检测设置，目标实例自动开始检测。

回到**设置 > 站点检测**页面，选择已绑定站点的实例，在其**操作**选项中，您可以执行以下操作：

- **暂停/启动检测**：如果您不希望在当前时间执行检测，您可以暂停检测；已暂停的检测，通过启动检测可以恢复。
- **编辑站点**：修改实例绑定的站点和检测频率信息。

> **说明**：
> 如果您修改了站点或首页地址，需要重新验证。

- **重新验证**：如果您的验证失效或在步骤6中选择了**稍后验证**，您可以重新验证对站点的管理权。
- **解除绑定**：如果您不希望继续向已绑定的站点提供检测服务，您可以解除绑定。

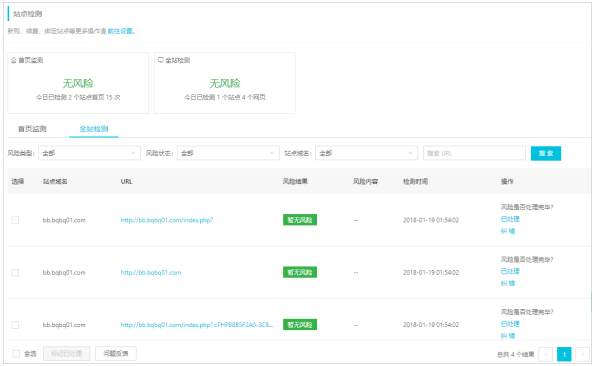
> **说明**：
> 解除绑定后，已购买的实例不会释放，但是您可以将其绑定到别的站点，为别的站点提供检测服务。

- **续费**：为目标实例续费，可以延长其使用时长。

查看检测结果

参照以下步骤，查看您的站点首页监测和全站检测的结果：

1. 登录云盾内容安全控制台。
2. 在左侧导航栏，选择**站点检测**。
3. 在**首页监测**和**全站检测**页签下，分别查看最近一次检测中发现的风险。



4. 通过提供的 URL，进一步查看并确认风险。

   - 消除风险后，单击**已处理**，完成处理。
   - 如果您对结果有异议，您可以单击**纠错**或**问题反馈**，通过表单将问题反馈给我们。在确认问题后，我们将在算法层面进行优化改进。

设置首页防篡改基准和重点监控URL

对绑定站点开启检测时，系统会抓取当前首页，作为判断首页是否被篡改的基准。若您更新过首页内容，建议您设置首页防篡改基准，重新抓取当前首页。如果您的网站内容很多，您担心在检测中会将重要的URL遗漏，您可以自定义重点监控URL，系统会优先检测您添加的URL。

参照以下步骤，为已启用站点检测的站点，设置首页防篡改基准和添加重点监控URL：

1. 登录云盾内容安全控制台。
2. 前往**设置 > 站点检测**页面。
   - **设置首页防篡改基准**
     1. 选择目标实例，在其操作选项中，选择**设置首页防篡改基准**。
     2. 确认当前首页基准。如果您想更换首页基准，单击**重新获取当前首页**，稍后即可查看到系统已重新抓取当前首页，作为首页基准。
   - **添加重点监控URL**
     1. 选择目标实例，在其操作选项中，选择**添加重点监控 URL**。
     2. 在对话框的输入区域，输入您想要添加的URL，每行一个URL，使用回车换行。最多支持添加5,000 条。
     3. 输入完成后，单击**提交**。

自定义词库

在站点检测时，如果您需要对特殊的词汇进行专门识别和防控，您可以自定义词库并将特殊关键词添加进来，进行黑名单防控。具体步骤如下：

1. 登录云盾内容安全控制台。
2. 前往**设置 > 站点检测**页面。
3. 单击**创建词库**。

   说明：
   最多支持创建10个词库。

4. 在**创建词库**对话框中，输入**词库名称**，并选择应用该词库的实例，然后单击**确定**。

   说明：
   在检测站点时，只有您选择的实例才会应用该词库。

5. 选择新创建的词库，在其**操作**选项中，选择**管理词库**。
6. 单击**新增关键词**，并按照页面提示输入或导入关键词。
7. 单击**确定**，完成添加。

自定义图库

在使用站点检测检查图片时，您可以将特定图片定义为白名单/黑名单图片，进行过滤/防控。具体步骤如下：

1. 登录云盾内容安全控制台。
2. 前往**设置 > 站点检测**页面。
3. 单击**创建图库**。

   说明：
   最多支持创建10个图库。

4. 在**创建图库**对话框中，完成相关配置，然后单击**确定**。配置说明如下：

配置	说明
图库名称	输入一个用于识别此图库的名称。
使用场景	选择智能监黄或暴恐涉政识别。

配置	说明
图库类型	选择黑名单或白名单。黑名单图库用于特殊防控不良图片，白名单图库会在检测中忽略并过滤您添加的图片。
选择实例	选择应用该图库的实例。 **说明：** 在检测站点时，只有您选择的实例才会应用该图库。

5. 选择新创建的图库，在其**操作**选项中，选择**管理图库**。
6. 单击**选择本地图片**，并上传本地图片至当前图库。

消息通知设置

云盾内容安全的默认消息推送每天触发1次。您可以设置消息接收方式、账号、和接收时间，也可以开启/关闭首页风险实时通知。具体步骤如下：

1. 登录云盾内容安全控制台。
2. 前往**设置 > 消息通知**页面。
3. 设置风险预警的**通知接收账户**（即接收邮箱地址和手机号码），勾选相应的**通知方式**（邮件、短信、站内信），和**定期推送时间**。

4. 考虑到首页风险的重要程度，您也可以设定是否开启首页风险实时通知。

   开启后，系统一旦检测出首页存在风险，会实时发送消息给您。如果多次检测到风险，为避免您被打扰，每天最多发送3次提醒。

3 OSS违规检测

内容安全通过人工智能技术鉴别OSS中的违规图片，帮助您减少90%以上审核人力，有效降低涉黄涉政风险。

观看以下视频，快速了解内容安全OSS违规检测功能。

前提条件

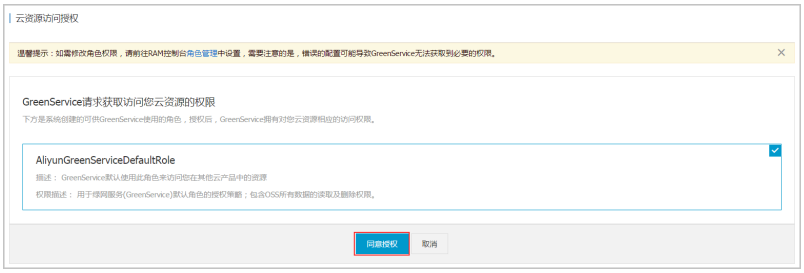
- OSS违规检测服务只向阿里云OSS用户提供，在使用前，请确保您已开通阿里云OSS服务。

  如果您还未开通OSS，请前往OSS 控制台，开通OSS服务。
- 您只有通过RAM授权云盾内容安全读取OSS Bucket的权限后，才能使用OSS违规检测。授权步骤如下：

  1. 登录云盾内容安全控制台。
  2. 在左侧导航栏，单击**OSS违规检测**。
  3. 根据页面提示，单击**授权**，授权云盾内容安全检测您的OSS图片。

     说明：

     如果您已授权内容安全检测 OSS，则此处没有该提示。
  4. 单击**同意授权**，完成访问授权。

- 如果您开启了OSS 防盗链，您必须在OSS 控制台将`https://yundun.console.aliyun.com`和`https://yundun.console.aliyun.com/?p=cts`增加至Refer白名单。
- OSS违规检测支持的endpoint包括：oss-cn-hangzhou.aliyuncs.com、oss-cn-shanghai.aliyuncs.com、oss-cn-qingdao.aliyuncs.com、oss-cn-beijing.aliyuncs.com、oss-cn-shenzhen.aliyuncs.com。

设置增量扫描

您可以对指定的 Bucket 中的指定图片（或视频）启用增量内容自动检测，并根据检测结果执行相应处理；也可以对 OSS 空间中的指定图片（或视频）进行存量内容一次性检测，等待扫描完成后，直接查看检测结果，并执行相应处理。

参照以下步骤，对指定 Bucket 设置增量检测计划：

1. 登录云盾内容安全控制台。
2. 前往设置 > **OSS** 违规检测。
3. 在**Bucket** 设置下，从左侧待选择框中勾选需要检测的Bucket，将其添加到右侧的已选择框中。

> 说明：
> 左侧待选择框中罗列了当前阿里云账号在OSS中的可检测的Bucket，上面的数字表示可选择的Bucket的数量（例如下图中的“15”）。

> 说明：
> 右侧已选择框中罗列了已选择的Bucket，上面的数字表示已选择的Bucket的数量（例如上图中的“9”）。单击已选择的Bucket右侧的漏斗图标可以设置过滤规则，仅扫描指定前缀（或目录下）的图片文件。例如，添加img/test_，则表示只扫描该Bucket中以`img/test_`为前缀的图片。

如果要扫描的图片文件在特定目录下，您可以在文件名前加上目录路径，以整体作为前缀。例如，您要扫描的文件在*img/201805*目录下，且前缀为*test_*，您可以添加*img/201805/test_*；如果您只想扫描*img/201805*目录下的图片，您可以添加*img/201805*。

> **说明：**
> 您最多可以为一个Bucket设置10条过滤规则。

4. 选择开启图片检测、视频检测，并完成相关配置。参数说明如下：

功能	配置	说明
图片检测	检测类型	勾选涉黄、涉政。
	自动冻结	开启自动冻结后，当检测分值高于指定分值时，自动冻结图片。开启后，需要分别设置涉黄自动冻结阈值和涉政自动冻结阈值。

		说明： 请慎重修改冻结阈值。默认冻结分数是100分，正常情况下不建议设置到99分以下，分数过低可能会导致正常图片被冻结，尤其是开启存量扫描时。冻结后的图片前台不可访问，帮助您防止风险外露。对于被冻结的照片，您可在**OSS违规检测**页面将其删除或者解冻。
	高级设置	开启/关闭限定每日图片扫描上限，并设置每日扫描量上限。 说明： 默认没有上限。如果设置了扫描上限，扫描数量超出限制后将会停止扫描，存在巨大违规图片外露的风险。常规情况下，不建议您设置扫描上限。
视频检测	检测类型	勾选涉黄、涉政。
	视频截帧设置	• 截帧频率：截帧频率越高，识别准确率越高。最低 60 秒 1 帧，最高 1 秒 1 帧。 • 单视频最大截帧数：设置超过多少帧以后不再截帧，作为上限保护参数。最少 100 帧，最多 9999 帧。 • 视频文件大小：设置仅扫描多少M以内大小的文件。检测视频大小不能超出 500M。
	自动冻结	开启后，违规视频将被自动冻结。勾选涉黄视频、涉政视频。

5. 勾选**我已同意《OSS 违规检测服务条款》**，并单击**保存**，保存后的设置即时生效。

设置成功后，系统会按照配置，自动对已选择的Bucket进行图片/视频检测。

执行存量扫描

您可以对指定Bucket中的指定图片（或视频）启用增量内容自动检测，并根据检测结果执行相应处理；也可以对OSS空间中的指定图片（或视频）进行存量内容一次性检测，等待扫描完成后，直接查看检测结果，并执行相应处理。

参照以下步骤，一次性扫描指定Bucket中的存量图片（和视频）：

1. 登录云盾内容安全控制台。

2. 前往**OSS 违规检测 > 存量扫描**。
3. 单击**开始扫描**，并确认。

4. 在**存量扫描设置**对话框完成相关设置。配置说明如下：

配置	说明
Bucket设置	选择要扫描的Bucket，并为已选择的Bucket设置扫描过滤规则。具体操作见设置增量扫描中步骤3说明。
时间范围	选择要检测的时间范围，该时间指图片上传到Bucket中的时间。
文件类型	勾选要检测的文件类型：图片、视频。
检测类型	勾选要检测的场景类型：涉黄、涉政。

配置	说明
	说明： 涉政和涉黄检测分开计费。例如，一张图片选择涉政和涉黄检测，既要计入涉政的计费，也会计入涉黄的计费。
其他设置	如视频截帧、自动冻结等，参考设置增量扫描中步骤4说明。

5. 单击**开始扫描**，并等待扫描完成，扫描结束后您会收到短信通知。

   在存量扫描过程中，您随时可以**终止扫描**。手动终止扫描后，扫描任务及配置会为您保留7天，供您在合适时间**继续扫描**。

检测结果处理

检测结束后，您可以在控制台查看、或按指定条件查询检测结果，并根据检测结果执行对应操作。

说明：
存量和增量检测结果页面类似，以下以增量检测页面为例进行说明。

1. 登录内容安全控制台。
2. 前往**OSS违规检测 > 增量扫描**页面，直接查看检测结果。

3. 设置查询条件，查询指定结果。

   您可以指定图片/视频、涉黄/涉政、分值范围、检测结果、Bucket、图片上传的时间范围来过滤结果数据，直接查看您希望看到的结果（图中标识①）。其中，

   - 分值范围：每一张图片下方会有违规检测服务检测出的风险分值，分值从0-100，分值越高，代表该图片是违规概率越大。
   - 检测结果：违规、疑似、正常。

4. 将鼠标放置在每一张图片/视频上，会显示与该图片/视频对应的操作项。您可以进行以下操作：

   - 单击违规并删除，可将图片/视频从内容安全控制台和OSS Bucket中一并删除（图中标识②）。
   - 单击正常并忽略，则忽略该检测结果。忽略后该图片/视频将不再在控制台展示，并不影响保存在OSS Bucket中的图片/视频（图中标识②）。
   - 单击图片/视频可将其选中，并支持多选。单击全选可以选中当前页的所有图片/视频，单击删除所选可将选中的图片/视频全部删除；单击忽略所选可将选中的图片/视频全部忽略（图中标识③）。
   - 若您设置了自动冻结功能，则还可以将已冻结的图片/视频解冻（图中标识③）。
   - 将鼠标放置在感叹号（！）上方可查看图片/视频的基本信息，包括创建时间、文件名和所在Bucket（图中标识④）。

4 内容检测API

4.1 授权访问MTS服务

在提交视频内容检测任务时，如果您选择通过OSS地址（oss://xxxx）上传视频URL的方式，则云盾内容安全对上传的OSS视频自动截帧。内容安全调用阿里云媒体处理服务（MTS）进行视频截帧，避免公网访问用户数据，最大限度降低流量费用。您必须授权MTS服务以内容安全的身份递交视频截帧任务。该操作通过阿里云访问控制中的角色管理功能实现，本文介绍了您需要完成的步骤。

背景信息

您只能在提交视频异步检测任务时，选择上传视频URL的方式进行配置。

关于访问控制服务的角色功能，请参考角色。

通过完成操作步骤，您将实现以下目的：

- 在您的阿里云账号下创建MTS服务角色，并指定由内容安全的阿里云账号扮演使用该角色。
- 授权所创建的MTS服务角色只读访问您的OSS空间。
- 在通过OSS地址上传视频URL时，按照格式要求拼接生成URL并上传。

这样，内容安全的阿里云账号将扮演所创建的MTS服务角色，调用自身MTS服务，访问您的OSS空间，获取视频内容并对其截帧。

操作步骤

1. 创建RAM角色。
   a) 登录*RAM*控制台。
   b) 前往角色管理页面，并单击页面右上角的新建角色。
   c) 选择角色类型为服务角色。

d) 选择受信服务为**MTS多媒体转码服务**。

e) 在**配置角色基本信息**页面，填写角色名称，并单击**创建**。

> **说明：**
>
> 该操作可能需要通过手机验证。

f) 创建成功后，回到**角色管理**页面，选择新创建的角色，单击**管理**。

g) 在**角色详情**页面，单击**编辑基本信息**。

h) 修改**策略内容**，将`"Service"`下的内容修改为`"1184847062244573@mts.aliyuncs.com"`，并单击**修改角色**。

该操作指定由内容安全的阿里云账号（UID：1184847062244573）扮演所创建的服务角色，调用其MTS服务。

2. 为服务角色授权。

a) 在角色管理页面，选择新创建的角色，单击授权。

b) 在可选授权策略名称下搜索授权策略**AliyunOSSReadOnlyAccess**，并将其添加到已选授权策略名称中。

该操作授权服务角色以只读权限访问您的阿里云账号下的OSS内容。

c) 单击确定。

3. 复制角色ARN（Aliyun Resource Name, 阿里云全局资源名称）。

a) 在角色管理页面，选择新创建的角色，单击管理。

b) 在角色详情页面，查看并复制其**Arn**。

4. 对要检测的OSS视频对象，按照以下格式拼接生成视频URL：`oss://arn@bucket.region/object`

   例如，假设您在深圳OSS的bucket *foo*上有视频对象*video/bar.mp4*需要检测，则拼接生成的URL为`oss://acs:ram::xxxxxxxxxxxxxxxx:role/mts-to-a@foo.cn-shenzhen/video/bar.mp4`（`xxxxxxxxxxxxxxxx`是您的16位阿里云ID。）

   > **说明：**
   > 目前支持的区域（region）包括：cn-hangzhou、cn-shanghai、cn-beijing、cn-shenzhen。

5. 提交视频检测任务时，上传拼接生成的URL作为检测对象。

4.2 数据统计

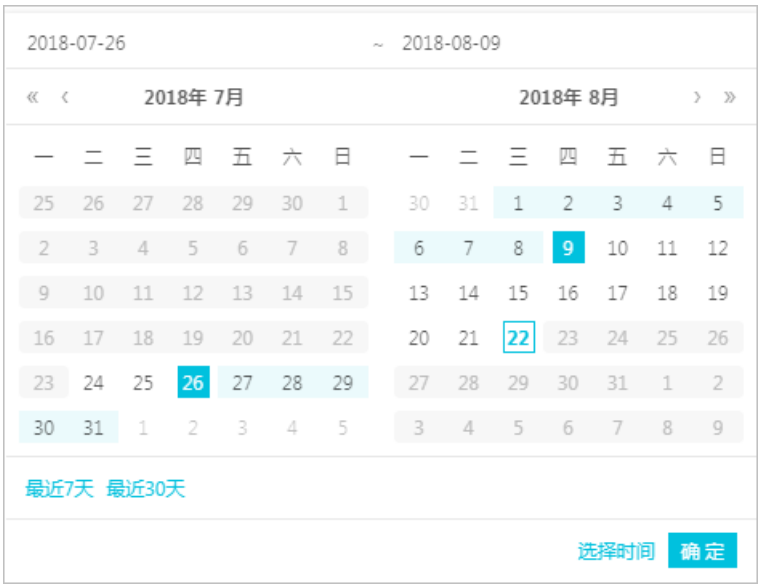
您可以在内容安全控制台查看内容检测API的调用数据统计。

背景信息

内容安全控制台汇总了内容检测API的调用统计数据，支持查询最近30天内调用图片、文本、视频检测接口的总次数，以及不同检测场景的结果中确认违规量、疑似违规量、和正常量。

操作步骤

1. 登录内容安全控制台。
2. 前往**内容检测API > 数据统计**页面。
3. 在**数据统计**页面，选择查询时间，并单击**查询**。支持查询的时间段为最近30内。您可以直接选择**最近7天**或**最近30天**；也可以分别选择一个起始日期和结束日期，查询在此时间段内的调用统计数据。



4. 选择要查询的检测接口类型：图片、文本、视频，分检测场景查看调用统计图表。

4.3 数据回流

为了持续改进检测效果，您可以通过数据标记回流，帮助内容安全API的检测模型针对性地学习您的审核标准，提升识别准确度。您可以使用阿里云自助审核平台和您自己的审核平台。

使用阿里云自助审核平台

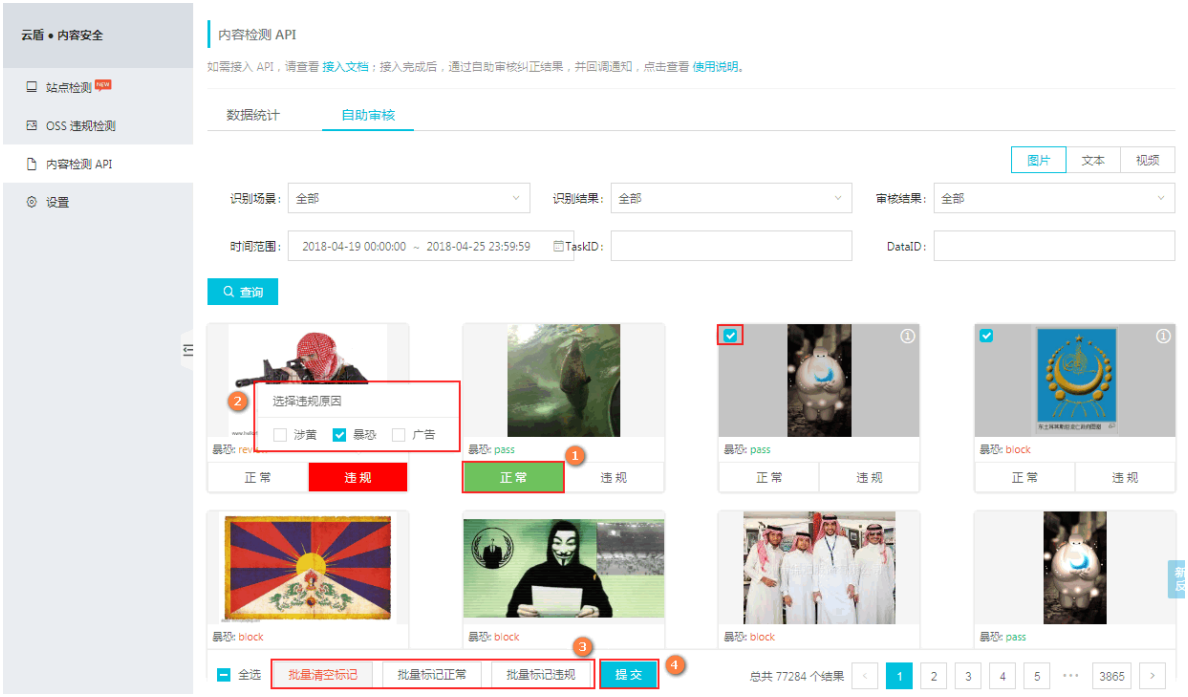
如果您没有自己的审核平台，建议您使用我们的自助审核平台进行审核和数据回流，审核结果将通过回调接口通知您。

标记样本

通过内容安全自助审核平台，您可以实时查看图片、文本和视频的识别结果，并对审核中发现的识别有误的样本进行标记，反馈给我们。如果您的审核人力有限，建议您重点审核被识别为违规（block）和疑似（review）的样本。

参照以下步骤，查看并标记识别结果：

1. 登录云盾内容安全控制台。
2. 前往**内容检测API > 自助审核**页面。
3. 按照以下方式，进行标记：



- 对于您认为正常，却被识别为违规（block）或者疑似（review）的样本，标记为**正常**（上图标识①）。
- 对于您认为需要管控，却被识别为疑似（review）或者正常（pass）的样本，标记为**违规**，并选择违规原因：**涉黄**、**暴恐**、**广告**（上图标识②）。

> **说明：**
> 支持勾选多张图片，进行批量处理，如**批量清空标记**、**批量标记正常**、**批量标记违规**（上图标识③）。

4. 标识完成后，单击**提交**（上图标识④）。被标记样本以及类似样本的检测结果将会按照您的标记结果实时纠正，同时会通过下文中的回调接口通知您。

结果通知

在对内容检测API检测结果进行审核时，您可以启动回调动作，并设置一个HTTP(s)接口。这样设置后，当您提交审核时，我们会通过该接口，将审核结果及系统检测的原始内容推送给您。

参照以下步骤，设置自主审核结果回调接口：

1. 登录云盾内容安全控制台。
2. 前往设置 > 内容检测**API**页面。
3. 在自主审核结果通知下，设置回调地址（即回调链接**callback**）。当您添加完回调链接时，系统会自动生成一个回调种子（**seed**）。

> **说明：**
> 最多只能添加一个回调链接（callback）。

关于回调通知参数（callback、seed），参见以下说明。

结果回调通知参数（callback、seed）

回调链接（callback）需支持POST方法，传输数据编码采用utf-8，并且支持表单参数checksum和content。系统将按以下描述的生成规则和格式设置checksum和content的值，调用您的callback接口，返回检测内容。

您的服务端接收到我们推送的结果后，返回的HTTP状态码为200时，表示推送成功，其他的HTTP状态码均视为您接收失败，我们将最多重复推送16次。

回调结果参数的生成规则

名称	类型	描述
checksum	String	由用户`uid + seed + content`拼成字符串，通过SHA256算法生成。用户UID即账号ID，您可在阿里

		云控制台上查询。为防篡改，您可以在获取到推送结果时，按此算法生成字符串，与checksum做一次校验。
content	String	JSON字符串格式，请自行解析反转成JSON对象。 content结果格式参见下文。

content结果格式

content包含以下两部分内容：

- 扫描结果（**scanResult**）：API调用返回结果中字段及描述。
- 审核结果（**auditResult**）：
  - **suggestion**：系统审核结果，取值：
    - pass：表示审核为正常。
    - block：表示审核为违规。
  - **labels**：审核为违规时的具体原因。

以下是一个content结果示例：

```
{
    "scanResult": {
        "code": 200,
        "msg": "OK",
        "taskId": "fdd25f95-4892-4d6b-aca9-7939bc6e9baa-1486198766695
",
        "url": "http://1.jpg",
        "results": [
            {
                "rate": 100,
                "scene": "porn",
                "suggestion": "block",
                "label": "porn"
            }
        ]
    },
    "auditReult": {
        "suggestion": "block",
        "labels": [
            "porn",
            "ad",
            "terrorism"
        ]
    }
```

```
}
```

使用您自己的审核平台

如果您有自己的审核平台，您可以直接对接反馈接口，将审核后认为识别有误的样本回流给我们。收到您的反馈后，我们会在下个版本的模型迭代中将您的反馈数据加入训练。关于反馈接口参数，参见具体接口文档。

说明：

模型训练需要积累足够的样本，可能无法立即生效。如果有需要，您可以开启自动加入自定义图库功能，实时纠正结果。

开启自动加入自定义图库的操作步骤如下：

1. 通过工单或者您的商务经理，联系内容安全运营人员，帮助您打开实时回流自定义图库的开关，选择要回流的场景。系统会帮助您在自定义图库中自动创建该场景的回流图库，分为黑名单和白名单。
2. 在反馈接口的**`label`**字段中，对于您认为正常的样本传`normal`，将该样本加入白名单；对于您认为违规的样本传任意字段（建议您使用porn、ad、terrorism等风险字段），将该样本加入黑名单。

说明：

在内容安全控制台，您也可以对回流图库像对其他自定义图库一样进行管理，只不过不能创建与删除回流图库。关于自定义图库的操作方法，请参考自定义图库。

4.4 自定义图库

在您出现突发性的管控需求，而模型更新时间较长，暂时无法满足需求时，内容安全提供了自定义图库功能，帮助您达到紧急止血的目的。

背景信息

自定义图库分为黑名单和白名单两种：

- 加入黑名单的样本以及类似样本，后续算法返回的**`suggestion`**都将为`block`。
- 加入白名单的样本以及类似样本，后续算法返回的**`suggestion`**都将为`pass`。

操作步骤

1. 登录云盾内容安全控制台。

2. 前往设置 > 内容检测API页面，并定位到自定义图库功能区。

3. 单击创建图库，并在创建图库对话框中完成以下操作：

a) 按照您的业务命名图库。

b) 选择使用场景：智能监黄、暴恐暴政识别、广告识别。

> **说明：**
> 该场景参数对应于API调用时通过**scenes**传入的参数。例如，假如该图库适用于鉴黄场景，则使用场景选择 智能鉴黄；那么，在您调用图片或者视频鉴黄服务时，都会默认启用该场景的黑白名单。

c) 选择图库类型：黑名单、白名单、疑似名单。

d) （可选）输入biztype。biztype属于高级功能，请根据需要进行设置。如果设置biztype，则使用biztype调用API时，带有该biztype的自定义图库才会生效。

e) 单击确定，完成创建。

4. 管理图库。

a) 回到图库列表中，选择目标图库，单击其**操作**列下的**管理图库**。

b) 在管理图库页面，您可以维护图库内的图片：

- 单击**选择本地图片**，上传图片到图库。

  说明：

  支持上传png、jpg、jpeg、bmp格式的图片文件，且单个文件需要小于5M，一次最多上传 20 张图片。

- 勾选不需要的图片，单击**批量删除**，删除图片。

5. 删除与修改图库。回到图库列表中，选择对应图库，单击其**操作**列下的**删除**或**修改**可以分别删除目标图库和修改目标图库的配置。

4.5 自定义词库

在您出现突发性的管控需求，而模型更新时间较长，暂时无法满足需求时，内容安全提供了自定义词库功能，帮助您达到紧急止血的目的。自定义图库支持将指定关键词加入到黑名单，加入黑名单的样本以及类似样本，后续算法返回的suggestion都将为block。

背景信息

操作步骤

1. 登录云盾内容安全控制台。
2. 前往**设置** > **内容检测API**页面，定位到**自定义词库**功能区。

3. 单击**创建词库**，在**创建词库**对话框中完成以下操作：

a) 选择使用**场景**：文本反垃圾、语音反垃圾、图片广告。

> **说明：**
> 该场景参数对应于API调用时通过scenes传入的参数。例如，假如该词库适用于文本反垃圾，则使用场景选择**文本反垃圾**；那么，在您调用文本（传入**scenes**）或文件检测服务（传入**textScenes**）时，都会默认启用该场景的黑名单。

b) 按照您的业务命名词库。

c) （可选）输入biztype。biztype属于高级功能，请根据需要进行设置。如果设置biztype，则使用biztype调用API时，带有该biztype的自定义图库才会生效。

d) 单击**确定**，完成创建。

4. 管理词库。

a) 回到词库列表中，选择目标词库，单击其**操作**列下的**管理词库**。

b) 在**管理词库**页面，您可以维护词库内的关键词：

- 单击**新增关键词**，按照页面提示在图库中增加要阻截的关键词。
- 勾选不需要的关键词，单击**批量删除**，删除关键词；也可以单击不需要的关键词下的**删除**，单独将其删除。

c) 已添加的关键词，您可以查看其**历史命中次数**，也可以使用搜索进行定位。

5. 删除与修改词库。回到词库列表中，选择对应词库，单击其**操作**列下的**删除**和**修改**可以分别删除目标词库和修改目标词库的配置。