

白皮书

在 AXIS Camera Station Pro 中进行自由文本搜索

1月 2025

概述

AXIS Camera Station Pro 预装了多种视频取证搜索工具。这些工具包括时间轴擦洗、数据搜索以及预分类目标和自由文本搜索的智能搜索。

通过自由文本搜索，您可以使用自己对任何移动目标的描述词语来进行搜索。可以使用多种描述性属性来自由创建详细的搜索过滤器，因此可以更快地找到相关片段。

自由文本搜索功能基于预先训练的开源基础模型所提供的文本与图像匹配，该模型已由安讯士针对监控用例进行了优化。可对一台摄像头进行搜索，也可对多台摄像头同时搜索。

将自由文本查询的数值表示与侦测到的移动目标图像的数值表示进行比较。搜索结果以缩略图的形式显示，包括摄像头名称、时间和日期，并按照与搜索查询的相关性排序。

通过自由文本搜索，我们利用人工智能提高取证搜索解决方案的准确性和效率，最终增强人的决策能力。为了遵守法律和伦理标准，搜索功能包括一个单独的安讯士开发的审核功能，可限制在搜索查询词中使用攻击性词语。所有搜索都会记录在案，管理员也能看到这些搜索，以便在出现不当使用时跟进并采取纠正措施。

目录

1	引言	4
2	背景：在 AXIS Camera Station Pro 中进行取证搜索	4
3	如何进行自由文本搜索？	4
4	构建搜索查询词	5
	4.1 提示指南	5
	4.2 查询词审核	5
	4.3 搜索查询词记录	6
5	文本与图像匹配	6
	5.1 专为监控任务优化的基础模型	6
	5.2 特征向量	6
6	流程概览	6
7	负责任地使用人工智能	8

1 引言

AXIS Camera Station Pro 中的自由文本搜索工具允许使用自己的词语而不是预定义的过滤器来搜索视频记录。

本白皮书概述了搜索方法的工作原理，对操作指南进行了说明，并介绍了为确保符合法律和伦理标准而设置的审核功能和查询词记录。

2 背景：在 AXIS Camera Station Pro 中进行取证搜索

AXIS Camera Station Pro 预装了多种视频取证搜索工具，包括时间轴擦洗、数据搜索以及预分类目标和自由文本搜索的智能搜索。

智能搜索功能采用安讯士设备生成的场景元数据。元数据包括移动目标的目标类型（人、车辆类型或未知目标），以及服装和车辆颜色、牌照、速度、位置和时间戳等属性。

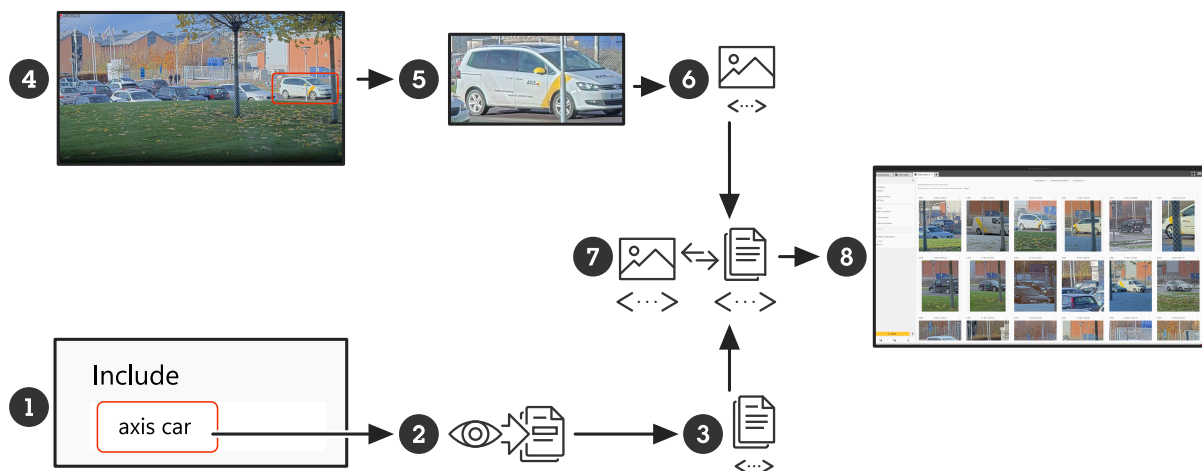
对于分析能力有限的设备，搜索功能基于设备中的移动侦测，以及在 AXIS Camera Station Pro 服务器上执行的目标分类。因此，AXIS Camera Station Pro 中的取证搜索是一种混合解决方案，即尽可能使用边缘设备的功能，并在必要时使用服务器的数据作为补充。

传统方法中，使用场景元数据进行搜索必须使用预定义的搜索过滤器。通过这些过滤器，可以从列表中选择固定的目标描述词，包括目标类型（如“车辆”）、车辆类型（如“轿车”（如适用））、颜色（如“蓝色”）等。新的自由文本搜索方法则允许创建自己的搜索过滤器。

虽然预分类搜索能提供精准结果，但它无法侦测到未预先定义的新型目标类型。为了解决这一限制，自由文本搜索提供了使用自己的词语进行搜索的自由和灵活性。您可以用自然语言和关联词语对任何移动目标进行更详细的描述，从而对搜索进行微调，获得更多结果。

3 如何进行自由文本搜索？

将自由文本查询的数值表示与侦测到的移动目标图像的数值表示进行比较。按最佳匹配度展示和排序文本与图像匹配的结果。搜索结果以缩略图的形式显示，包括摄像头名称、时间和日期，并按照与搜索查询的相关性排序。



自由文本搜索流程简化概览。请注意，第 4 至第 6 步会持续不断地进行，即便不搜索时也在创建所有侦测到的移动目标的特征向量。

- 1 输入自由文本搜索查询词。
- 2 审核模块可阻止使用毒性和不道德的词语。
- 3 基础模型创建搜索查询的数值表示（特征向量）。
- 4 摄像头侦测场景中的移动。
- 5 摄像头选择一张裁剪图像来表示移动目标。
- 6 基础模型在分析目标的形状、图案、颜色等之后，创建目标的特征向量。
- 7 将两类特征向量进行比较。
- 8 比较结果按最佳匹配度排序，并以缩略图的形式显示。

可对一台摄像头进行自由文本搜索，也可对多台摄像头同时搜索。

要缩小自由文本搜索的范围，可以将其与其他智能搜索功能（如相似性搜索或基于时间的搜索）相结合，使用一种又一种搜索类型。

4 构建搜索查询词

您可以搜索任何移动目标和任何类型的车辆。要获得最佳搜索结果，请遵循操作指南。

请注意，只能搜索移动的目标。大多数情况下无法搜索静止的目标。

搜索短语须经审核并被记录在案，以防止不道德的搜索行为。

4.1 提示指南

- 像描述图像一样描述情况。该模型接收的是静态图像，因此搜索动作（如摔倒、奔跑或偷窃）可能比较困难，因为这些动作需要更多的上下文信息。
- 使用几个关键描述词来描述目标：“一个穿红毛衣的人”或“一辆黄色皮卡车”。与其他多模态模型一样，自由文本搜索模型在处理目标和颜色等描述性信息时表现良好，但在处理计数（“三个人”）、俚语或情感提示（“一脸怒气的男人”）时效果较差。目标描述不应主观、过于模糊，也不应包含过于具体的细节。
- 用“并”组合多个目标属性：“戴红帽并背背包的人”。
- 描述文本、文字标识或品牌名称：“带有‘安讯士’文字的面包车”。
- 不要把重点放在描述环境上。模型处理的是目标的裁剪图像，即模型可能无法看到目标的周围环境。因此，宽泛的场景或环境描述词（如“城市”、“都市”、“公园”、“花园”、“湖泊”或“海滩”）可能效果不佳。
- 如果对结果不满意，可以用其他短语再次搜索。
- 自由文本提示功能仅支持英语。

4.2 查询词审核

我们根据常见做法实施了查询词审核，以确保有效过滤。审核模型是一种自然语言处理模型，用于检查查询词，限制攻击性措辞。它可以检查整个文本字符串是否存在有害、不当或毒性内容。此外，我们还通过专有措施增强了这些功能，包括定制禁止搜索类别和词语列表。当查询词中包含该列表中的词语或短语时，我们会拒绝搜索，以维护安全的搜索环境，确保搜索结果符合伦理标准。如果不同意某个词语被屏蔽或想建议屏蔽某个词语，您可以向安讯士提供匿名用户反馈。

4.3 搜索查询词记录

AXIS Camera Station Pro 可对用户操作进行审查追踪。审查追踪不仅跟踪具体操作和用户身份，还会保留操作中使用的任何数据。也就是说，所有用户搜索（包括搜索提示）都会被记录在案。管理员可以利用日志识别用户的不当搜索行为，标记不道德的搜索提示，并采取纠正措施。

请注意，视频数据不会与安讯士共享。您的数据将保留在您的服务器上。

5 文本与图像匹配

使用自由文本查询可搜索视频元数据，这极大扩展了搜索功能，从预定义的属性列表扩展到几乎无限的搜索标准。在 AXIS Camera Station Pro 中，该功能基于一个开源基础模型，该模型利用数十亿图像与文本对进行训练，并由安讯士针对监控用例进行微调，从而提高性能。

5.1 专为监控任务优化的基础模型

基础模型是一种通过大规模文本和图像对数据集训练的模型，能够将文本输入转换为图像输出。这是一种零样本模型，可将文本与相关图像进行匹配。零样本模型是一种人工智能 (AI) 模型，可以在没有预训练数据的情况下对目标或概念进行识别和分类。换言之，该模型可以在之前未见过任何任务实例的情况下执行任务。这种能力对于确保我们提供最佳的自然语言与图像匹配性能至关重要。

该模型通过大量文本和图像组合进行训练，并在一个拥有超过 25 亿个参数的神经网络上运行。安讯士使用独有的训练材料进一步微调了该模型，提升了它在处理典型监控摄像头视角和目标时的图像解析能力。也就是说，我们针对监控用例对该模型进行了优化。

5.2 特征向量

进行自由文本搜索时，基础模型会创建搜索查询的特征向量。

基础模型还会不断生成场景中每个被跟踪目标的描述，并创建特征向量来表示这些目标。每个目标只用一个特征向量表示，该向量存储在我们的数据库中。由于特征向量已在数据库中预先计算并随时可用，因此搜索速度很快。

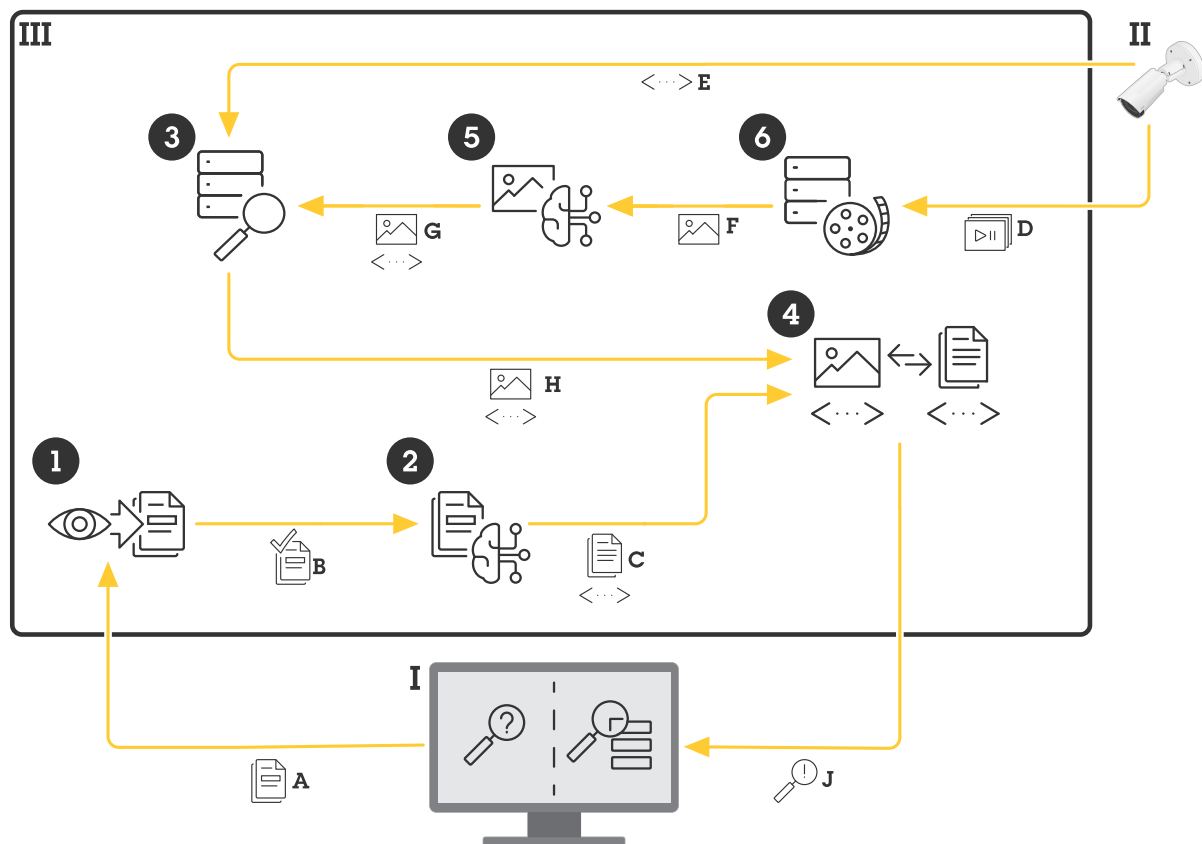
两类特征向量都会被输入到向量比较引擎中，以确定搜索查询词与表示侦测到的目标的所有可用特征向量之间的相似距离。

特征向量是文本或图像的数值表示。因此，人或目标的特征向量只是人或目标外观的抽象表征。特征向量不包含任何人类可解读的特征信息，如头发或衣服颜色，这些特征信息可以明确映射到特定的人或用于确认身份。特征向量只能用于与其他特征向量进行比较。

6 流程概览

流程概览图显示了主要流程步骤，包括各步骤在什么位置进行，以及每个步骤生成哪类数据。

请注意，图中上部的环路包括摄像头 (II)、录像存储 (6)、基础模型 (5) 和搜索数据库 (3)，这是一个持续不断地创建所有侦测到的移动目标的特征向量的流程，而不是只在搜索时才创建。



自由文本搜索流程的主要位置 (I-III)

I AXIS Camera Station Pro 客户端

II 摄像头

III AXIS Camera Station Pro 服务器

主要流程步骤 (1-6)

1 搜索查询词审核

2 基础模型 (文本)

3 搜索数据库

4 向量比较

5 基础模型 (图像)

6 录像存储

数据类型或结果 (A-J)

A 文本字符串

B 文本字符串

C 特征向量 (文本)

D 视频

E 元数据

F 图像

G 特征向量 (图像)

H 特征向量 (图像)

J 搜索结果

(I) **AXIS Camera Station Pro 客户端**：您可以在客户端输入搜索查询词，获得排序的搜索结果

(II) **摄像头**：自由文本搜索适用于使用 AXIS OS 5.51 或更高版本的 Axis 摄像头，摄像头版本越高，搜索结果越佳。旧版本设备只根据移动侦测提供粒度较小的元数据。新版本设备可生成 AXIS Scene Metadata，其中包括目标分类。利用摄像头的移动目标侦测和跟踪功能，可以为每个侦测到的目标找到一张具有代表性的图像，从而减少服务器上需要分析的图像数量。

(III) **AXIS Camera Station Pro 服务器**：摄像头的所有元数据和视频数据都在服务器中处理并存储在服务器中。进行自由文本搜索之前，服务器必须（针对每个侦测到的移动目标）对视频进行解码，并提取侦测到的目标的图像。然后，基础模型对该图像进行处理，创建特征向量。这些操作对处理能力的要求很高，因此为了提高性能，需要将特征向量保存到数据库中，以便日后快速启用搜索功能。如果您的服务器有剩余容量，我们强烈建议您对最重要摄像头所拍视频启用后台处理功能，这将大大加快搜索速度。

(1) **搜索查询词审核**：审核模型会检查查询词以限制攻击性内容。

(2) **基础模型（文本）**：基础模型将创建经审核的搜索查询文本字符串的数值表示（特征向量）。

(3) **搜索数据库**：搜索数据库存有 AXIS Scene Metadata 的完整元数据或服务器创建的元数据，包括带有属性、时间、位置和特征向量的目标分类数据。

(4) **向量比较**：将搜索查询文本字符串的特征向量表示与视频中侦测到的目标图像的特征向量表示进行比较。

(5) **基础模型（图像）**：基础模型为录制的视频中的每个目标轨迹创建数值表示（特征向量）。这是一个持续不断的流程，即使不搜索也在创建。

(6) **录像存储**：这里存储着摄像头所拍的视频，也是基础模型获取图像之处。

7 负责任地使用人工智能

通过自由文本搜索，我们利用人工智能提高取证搜索解决方案的准确性和效率，最终增强人的决策能力。

责任和问责是安讯士人工智能方法的基础。这需要确保我们创建的人工智能系统遵循伦理原则、遵守法律并能有效管理风险。安讯士提供的工具让我们的客户对完善的系统运行充满信心。AXIS Camera Station Pro 的自由文本搜索功能包括一个微调文本分类模型，用于文本提示审核。我们开发这一模型是为了审核搜索查询词，有助于阻止不道德的使用。

自由文本搜索每周一次连接到安讯士云服务，检查人工智能模型是否需要更新，确保符合新法规或要求。如果连接失败，将无法进行自由文本搜索，直到重新建立连接。

为了进一步遵守人工智能应用中的法律和伦理标准，我们的产品提供基于用户身份验证凭证和访问权限的访问控制。这样，我们的客户就可以强制用户遵守操作策略。

关于安讯士 (Axis Communications)

安讯士通过打造各种解决方案，提高安全水平和企业效益，旨在创建一个高度智能、更加安全的世界。作为一家网络技术公司和行业领导者，安讯士致力于推出视频监控、访问控制、内部通信和音频系统解决方案。安讯士通过智能分析应用程序增强解决方案，并提供高质量培训支持。

安讯士在全球50多个国家和地区设有办事机构，拥有超过4,000名尽职的员工，并与遍布世界各地的技术和系统集成合作伙伴携手并进，为客户带来高价值的解决方案。安讯士创立于1984年，总部位于瑞典隆德