

视频分析中的人工智能 (AI)

依托机器学习和深度学习的分析技术

3月 2021

目录

1	概述	3
2	引言	4
3	AI、机器学习与深度学习	4
	3.1 机器学习	4
	3.2 深度学习	5
	3.3 经典机器学习与深度学习	6
4	机器学习的开发阶段	6
	4.1 数据收集和数据注释	7
	4.2 训练	7
	4.3 测试	8
	4.4 部署	8
5	前端分析	8
6	硬件加速	9
7	AI仍然处在初期发展阶段	9
8	保障分析性能的注意事项	9
	8.1 图像可用性	10
	8.2 侦测距离	10
	8.3 报警与录像设置	10
	8.4 维护	11
9	隐私与个人信息保护	12
10	附录	13
	10.1 神经网络	13
	10.2 卷积神经网络 (CNN)	13

1 概述

基于AI的视频分析是视频监控行业讨论很多的话题之一。某些应用程序能够显著提升数据分析速度，自动开展重复性任务。但现今，AI解决方案无法取代人类操作员的经验和决策制定能力。它的切实优势在于组合：利用AI解决方案，改善和提高人工效率。

AI概念包含机器学习算法和深度学习算法。这两种类型都能够使用大量采样数据（*训练数据*）自动构建数学模型，以便能够在无需专门编程的情况下计算结果。AI算法的开发通过迭代过程实现，其中，在得到所需的质量水平之前，要反复执行如下操作循环：收集训练数据、对训练数据加标记、使用加标记的数据训练算法以及测试经训练的算法。此后，可随时将算法用于分析应用程序中，此应用程序在购买后即可部署在监控点。这时，所有训练都已完成，应用程序将不再进行任何新的学习。

基于AI的视频分析的典型任务是，以视觉方式侦测视频流中的人和车辆，并对它们加以区分。*机器学习*算法已学习对这些目标予以定义的视觉特征的组合。*深度学习*算法是进一步优化，在经过相应训练后，能够侦测复杂得多的目标。但在使用最终的应用程序时，它需要的开发量和训练量以及计算资源也多得多。因此，应依据明确的监控需求，考虑专门的、经优化的机器学习算法是否已经足够。

摄像机的算法开发和不断提升的处理能力使得能够直接在摄像机上（基于前端）运行先进的基于AI的视频分析工具，而不必在服务器上（基于服务器）执行计算。这样就能够改善实时功能性，因为应用程序能够直接访问未压缩的视频材料。相比CPU或GPU（图形处理单元），利用摄像机中的专用硬件加速器，如MLPU（机器学习处理单元）和DLPU（深度学习处理单元），能够更省电地实现前端分析。

在安装基于AI的视频分析应用程序之前，必须仔细研读并遵守制造商基于已知前提条件和限制要求所提出的建议。每套监控安装都具有唯一性，应针对每个监控点评估应用程序的性能。如果质量低于预期，应开展全面调查，而不是仅着眼于分析应用程序本身。视频分析性能取决于多方面因素，涉及摄像机硬件、摄像机配置、视频质量、场景动态和照明等。在许多情况下，了解这些因素的影响并进行相应的优化有助于提升系统的视频分析性能。

随着AI在监控领域的应用日益广泛，必须严谨评估这项技术的使用时机和场合，从而在运行效率优势与新应用场合之间合理平衡。

2 引言

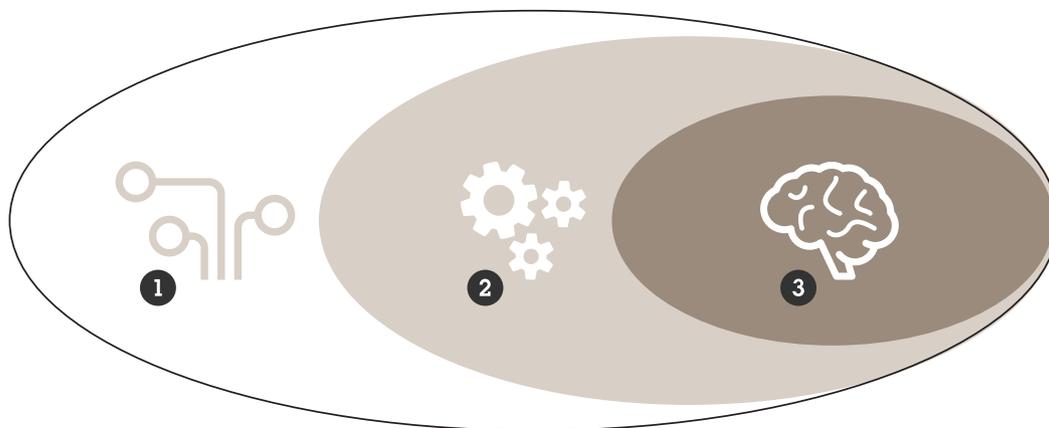
自首台计算机面世以来，AI（人工智能）就被开发了出来，有关它的争论也从未停止。虽然目前尚无最具变革性的代表产品，但基于AI的技术当今已在语音识别、搜索引擎、虚拟助手等应用中广泛用于开展明确的任务。AI在医疗保健领域的应用也日益广泛，它可在（比如）X光诊断和视网膜扫描分析中提供宝贵的资源。

基于AI的视频分析是视频监控行业讨论很多的话题之一，相关的预期也较高。市面上已有应用产品采用AI算法来成功实现了数据分析速度的提升以及重复性任务的自动开展。但就较广泛的监控情景而言，在现今以及在不久的将来，AI应仅被视作一种元素，尤其是在构建正确解决方案的过程中，更应如此。

本白皮书介绍了有关机器学习和深度学习算法的技术背景，以及如何开发这些算法并将它们应用于视频分析。其中还简要介绍了AI加速硬件，以及较之于服务器，在前端运行基于AI的分析有哪些优缺点。本白皮书还介绍了，在广泛兼顾多方面因素的情况下，如何优化基于AI的视频分析性能的先决条件。

3 AI、机器学习与深度学习

人工智能(AI)是一个广泛概念，它与能够解决复杂任务同时又有着貌似智能特性的机器相关。深度学习和机器学习是AI的细分类。



- 1 人工智能
- 2 机器学习
- 3 深度学习

3.1 机器学习

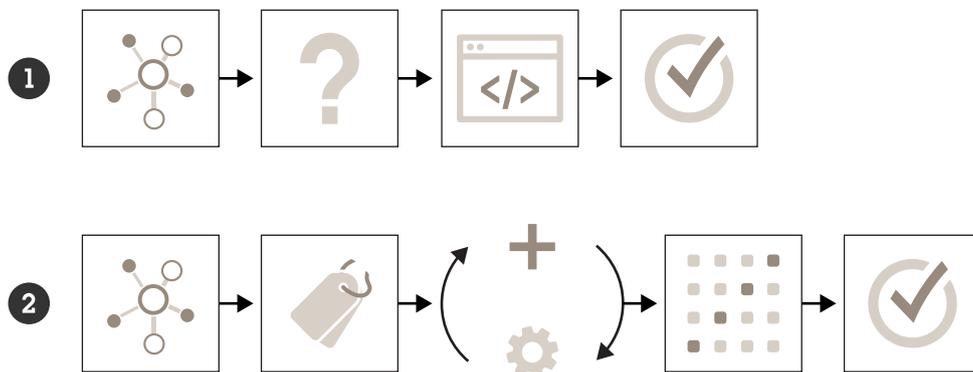
机器学习是AI的一个细分类，它使用统计学习算法构建在训练期间能够在无需显式编程的情况下自动学习并改进的系统。

这里，我们在*机器视觉*的情景下，对传统编程与机器学习进行了区分 – 后者通过分析图像或视频，使计算机理解场景中正在发生什么。

经传统编程的计算机视觉基于计算图像*特征*的方法，比如，查找明显边缘和拐角点的计算机程序。这些特征需要由算法开发人员手动定义，这些开发人员知道图像数据中哪些是重要数据。开发人员然后组合这些特征，以供算法用于推断在场景中所找到的信息。

机器学习算法使用大量采样数据（*训练数据*）自动构建数学模型，以便能够在无需专门编程的情况下，通过计算结果做出判定。这些特征仍然是人为设计的，但如何组合这些特征由算法自身通过对大量加标记（或*加注释*）的训练数据的接触来学习。在本文中，我们将这种在学习组合中使用人为设计的特征的技术称为*经典机器学习*。

换言之，对于机器学习应用，我们需要训练计算机来获取我们希望的程序。数据在收集后由人来加注释，有时还要辅以服务器计算机的预先注释。结果被送入系统中，这个过程会一直持续，直到应用程序所获得的学习足以实现我们想要的侦测，比如侦测特定类型的车辆。经训练的模型就变成了程序。请注意，当程序执行结束时，系统不再进行新的学习。



- 1 传统编程：
收集数据。定义程序条件。对程序编码（人工）。完成。
- 2 机器学习：
收集数据。标记数据。模型历经迭代训练过程。最终的模型变成程序。完成。

在构建计算机视觉程序时，AI相比传统编程的优势在于，能够处理大量数据。计算机能够全神贯注处理数以千计的图像，而人类程序员在一段时间后就会疲倦且无法专注。因此，AI能够大大提升应用程序的准确度。但应用程序越复杂，机器得到预期结果的难度就越大。

3.2 深度学习

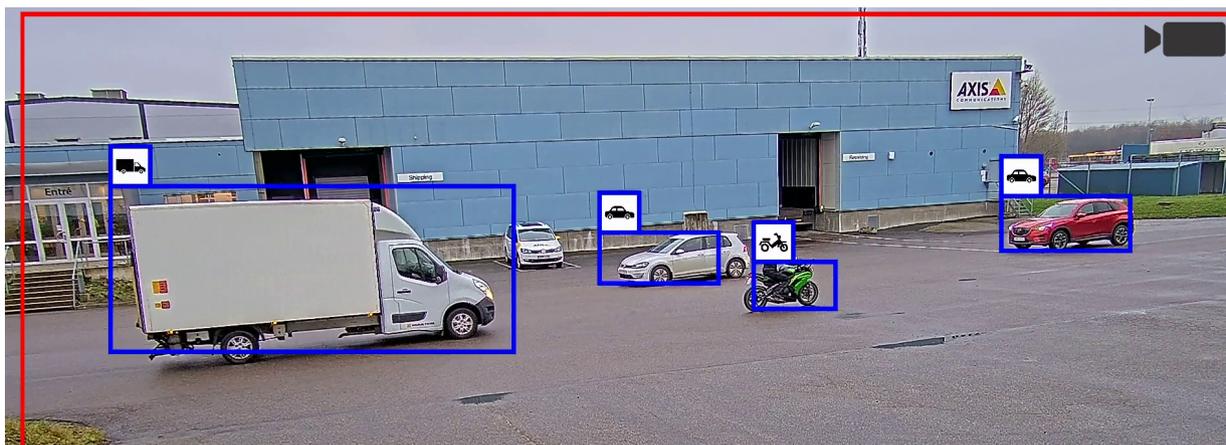
深度学习是机器学习的优化版本，在这种模式中，以数据驱动的方式，在用于生成输出的深度规则结构中，学习特征提取以及如何组合这些特征。这种算法能够自动定义要在训练数据中寻找的特征。它能够学习链式特征组合的非常深层次的结构。

深度学习中使用的算法的核心基于神经元的工作方式、以及大脑如何使用这些神经元通过以链式法则的深层次结构（或*网络*）组合神经元输出来形成上层知识。大脑就是一套系统，其中，组合本身也是由神经元形成的，这就消除了特征提取与特征组合之间的区别，使得它们在某种意义上是相同的。这些结构被研究人员仿真成*人工神经网络*，这是深度学习中广泛使用的算法类型。有关神经网络简述，请参见本文档的附录。

利用深度学习算法，能够构建复杂的视觉侦测器，并自动训练这些设备来侦测复杂的目标，灵活适应不同比例、旋转和其他型式。

这种灵活性的原因在于，相比经典机器学习系统，深度学习系统能够从大量的数据以及广泛的数据形式中进行学习。在大多数情况下，它们将大大优于人为设计的机器视觉算法。这

使得深度学习尤其适用于解决复杂问题，诸如图像分类、语言处理和目标侦测，因为在这些情形中，专业人员无法轻松形成特征组合。



基于深度学习的目标侦测可以对复杂目标分类。例如，分析应用程序不仅能够侦测车辆，而且还能够对车辆类型分类。

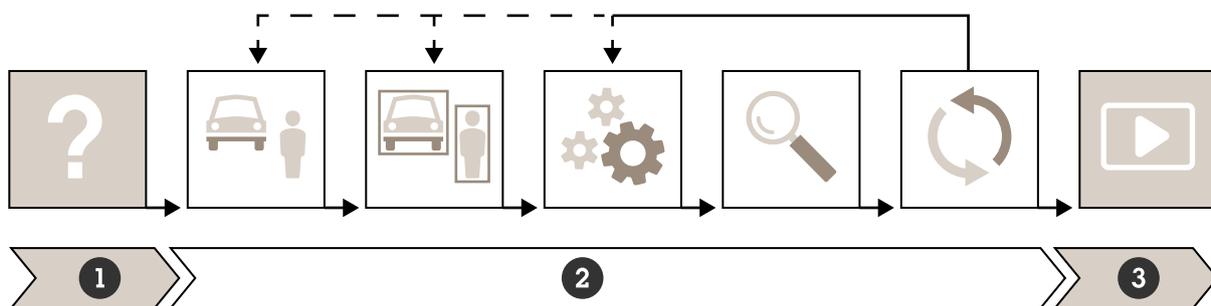
3.3 经典机器学习与深度学习

虽然这是两种相似的算法，但深度学习算法使用的学习特征组合通常比经典机器学习多得多。这就意味着，基于深度学习的分析能够更灵活，并且在经过相应训练后，能够学习执行复杂得多的任务。

但对于特定监控分析而言，专门的、经优化的经典机器学习算法可能已经足够。在有效监控范围内，它提供的结果与深度学习算法的结果不相上下，但需要的数学运算却更少，因此使用成本可能更低、电源占用可能更小。此外，它需要的训练数据也少得多，这就大大降低了开发工作量。

4 机器学习的开发阶段

机器学习算法的开发遵循一系列步骤和迭代（大致如下所示），然后才能够部署最终的分析应用程序。分析应用程序的核心是一种或多种算法，比如目标侦测器。如果是基于深度学习的应用程序，则算法的核心是深度学习模型。



1 准备：定义应用程序的用途。

- 2 **训练**：收集训练数据。对数据加注释。训练模型。测试模型。如果质量与预期不符，则在迭代改进循环中，再次执行先前的步骤。
- 3 **部署**：安装并使用最终的应用程序。

4.1 数据收集和数据注释

如要开发基于AI的分析应用程序，需要收集大量数据。在视频监控领域中，这通常包括人和车辆或其他相关目标的图像和视频片段。为了让机器或计算机能够识别这些数据，必须执行数据注释过程，期间将对相关目标分类并加标记。数据注释主要由人工完成，任务繁重。预处理数据需要涵盖种类足够广泛且与分析应用程序预期的应用情境相关的样本。

4.2 训练

训练或学习是指，为模块送入加注释的数据，并使用训练框架反复修改和改善模型，直到获得所需的质量。换言之，通过优化模型来解决预定义任务。训练方法主要有三种，可任选其一。



- 1 **监督学习**：模型学习执行准确的预测
- 2 **无监督学习**：模型学习识别数据集群
- 3 **强化学习**：模型从错误中学习

4.2.1 监督学习

监督学习是当今机器学习中常使用的方法。它可以被表述为示例学习。训练数据被明确加注释，这就意味着，输入数据已经与所需的输出结果配对。

监督学习所需的注释数据量通常很大，经训练的算法的性能直接取决于训练数据的质量。最重要的质量方面是，使用代表来自实际部署情景的所有潜在输入数据的数据集。对于目标侦测器，开发人员必须确保以种类广泛的图像以及不同的目标实例、方向、比例、光照情形、背景和干扰来训练算法。只有在训练数据能够表征所计划的应用场合时，最终分析应用程序才能够在处理于训练阶段中未见到过的新数据时，也执行准确的预测。

4.2.2 无监督学习

无监督学习利用算法来分析和分组未标记的数据集。它不是监控行业的常用训练方法，因为其模型需要大量校准和测试，而质量可能仍然无法预测。

数据集必须对于分析应用程序而言具有相关性，但不必明确加标记或标签。不需要手动加注释，但训练所需的图像或视频量必须大大增加达数倍之多。在训练阶段中，待训练模型在训练框架的支持下，识别数据集中的公共特征。这样，在部署阶段中，就能够根据样式对数据分组，同时还能够检测出不匹配学习组的异常。

4.2.3 强化学习

强化学习被用在机器人、工业自动化、商业策略规划等应用中，但由于需要大量的反馈，这种方法在监控领域的应用目前依然有限。强化学习涉及采取适当措施最大程度提高在特定应用情形下的潜在回报，在模型做出正确选择时，这种回报会进一步提升。算法不使用数据/标记对进行训练，而是通过与环境交互并同时衡量回报的方式，来检测决策，从而得以优化。算法的目的是学习策略，以便开展有助于实现更大化回报的措施。

4.3 测试

一旦对模型进行训练，便需要对其进行全面测试。这个步骤通常包含一个自动开展的部分，同时还辅以面向真实部署情形的广泛测试。

在这个自动开展的部分中，以模型在训练期间未见到过的新数据集对应用程序执行基准测试。如果这些基准测试的结果与预期不符，则会重新开始以下过程：收集新训练数据、加注释或优化注释、以及对模型进行再训练。

在达到期望的质量水平后，开始现场测试。在这个测试中，应用程序被应用到真实场景中。具体的测试量和型式取决于应用范围。应用范围越小，需要测试的型式越少。应用范围越大，需要的测试越多。

然后再次比较并评估结果。这个步骤随后可能再次导致重启上述过程。另一个潜在结果可能是，需定义前提条件，说明不应使用或仅部分推荐使用该应用程序的已知场景。

4.4 部署

部署阶段也被称为推断或预测阶段。*推断或预测*是执行经训练的机器学习模型的过程。算法使用在训练阶段中所学习的信息来得到预期的结果。在监控分析情景中，推断阶段是在监控系统上运行的应用程序监控真实场景的阶段。

在对音频或视频输入数据执行基于机器学习的算法时，如要实现实时性能，通常需要特定硬件加速。

5 前端分析

高性能视频分析曾经一度基于服务器，因为他们对电源能力和冷却能力的要求较高，而摄像机无法提供这样的能力。但近些年，前端设备的算法开发和不断提升的处理能力已使得能够直接在前端运行先进的基于AI的视频分析工具。

前端分析应用程序的优点非常明显：它们能够以很低的延迟访问未压缩的视频材料，从而确保实时应用，同时还能够避免因将数据转移到云端进行计算而造成的额外成本和复杂度。由于监控系统所需的服务器资源较少，前端分析涉及的硬件和部署成本也较低。

在某些应用中，也可以将前端处理与服务器处理组合使用，让摄像机负责预处理，随后让服务器负责深化处理。这样的混合系统通过对若干摄像机流的处理，能够有助于以成本效益好的方式扩展分析应用程序。

6 硬件加速

虽然通常可以在多种类型的平台上运行分析应用程序，但在处理能力有限的情况下，使用专门的硬件加速工具能够实现更高的性能。硬件加速器有助于更省电的分析应用程序。此外，还可以视情况辅以服务器和云计算资源。

- **GPU（图形处理单元）**。GPU主要用于图形处理应用，但也可用于服务器和云平台上的AI加速。虽然有时也用在嵌入式系统（前端）中，但从省电效率上讲，GPU不是机器学习推断任务的理想选择。
- **MLPU（机器学习处理单元）**。MLPU能够加速特定经典机器学习算法的推断，从而以省电的处理效率解决计算机视觉任务。它设计用于对种类有限的若干目标（比如，人和车辆）同时进行实时目标侦测。
- **DLPU（深度学习处理单元）**。内置有DLPU的摄像机能够加速通用深度学习算法推断，同时兼顾高效省电处理，实现更细化的目标分类。

7 AI仍然处在初期发展阶段

首先，我们不禁要将AI解决方案的潜力与真人的能力进行一下比较。负责视频监控的人类操作员只能在较短时间内保持全神贯注，而计算机能够不知疲倦地快速持续处理大量数据。但如果认为AI解决方案将取代人类操作员，这是一种误解。前者的切实优势在于基于现实的组合：利用AI解决方案，改善和提高人类操作员的效率。

机器学习或深度学习解决方案通常被描述为，具有自动学习或通过经验实现改进的能力。但当前市面上的AI系统还不会在部署后自动学习新技能，也不会记住已发生的特定事件。为了提升系统性能，在监督学习会话期间，需要以更好更准确的数据进行再训练。无监督学习通常需要大量数据，才能生成数据集群，因此不适用于视频监控应用。当今，它主要用于分析大数据集，以查找异常，例如，在财务交易中。在视频监控领域中，大多数被视为“自学”的方法都基于统计分析，而不是基于对深度学习模型的实际再训练。

就监控目的而言，人类经验仍然完胜许多基于AI的分析应用。这尤其体现在需要执行普通的任务以及情景理解很关键的那些场合中。基于机器学习的应用程序在经过专门的训练后，也许能够成功侦测“正在奔跑的人”，但与能够将数据与情景相结合的人类不同，应用程序无法理解这个人为什么在奔跑 – 是要赶公共汽车，还是在逃避附近警察的追捕？尽管有企业承诺将AI应用到自己的分析应用中以实施监控，但应用程序仍无法像人类那样做到对视频内容的理解。

出于同样的原因，基于AI的分析应用还可能触发假报警或不发出报警。这通常可能发生在涉及大量运动的复杂环境中。但这也可能是因为，比如，拿着大物件的人 – 人体特征遭到遮挡，应用程序无法有效捕捉，从而难以正确分辨。

现今，基于AI的分析应作为辅助手段使用，例如，用于先大致判断事件的相关性，然后再向人类操作员发出警报以做出响应决策。这样，AI就是用于能力扩展的，而人类操作员负责评估潜在的事件。

8 保障分析性能的注意事项

如要达到AI分析应用程序的质量预期，建议仔细研究并理解相关前提条件和限制要求，这些通常在应用程序的文档中已列出。

每套监控安装都应针对监控点评估应用程序的性能。如果质量未达到预期或期望的水平，强烈建议不要仅着眼于调查应用程序本身。应开展全面调查，因为分析应用程序的性能取决于

许多因素，而其中大多数因素都可以在了解它们的影响后加以优化。这些因素包括，例如，摄像机硬件、视频质量、场景动态、照明水平、以及摄像机的配置、位置和朝向。

8.1 图像可用性

图像质量通常被认为是取决于摄像机的高分辨率和高感光度。虽然这些因素的重要性毋庸置疑，但确实也有其他因素同样会影响到图像或视频的实际可用性。比如，如果监控场景夜间照明不良，如果摄像机朝向改变，或者如果系统连接断开，即便是昂贵的监控摄像机提供的出色的视频流，也可能变得毫无用处。

在部署之前，应仔细斟酌摄像机的安置。为了使视频分析工具达到预期性能，摄像机的安置需要确保能够清晰看到目标场景，视野中无障碍物。

图像可用性还可能取决于应用场合。人眼看起来很好的视频可能并不代表视频分析应用程序的更好性能表现。事实上，在使用视频分析时，不建议使用通常用于面向人眼增强视频视效的许多图像处理方法。其中可能包括（比如）降噪方法、宽动态范围方法、或自动曝光算法。

当今的视频摄像机通常随附有集成式红外照明装置，这让它们能够在漆黑环境中工作。优点在于，摄像机能够安置在光照差的地方，降低了对额外安装照明装置的需求。但如果安置点有大雨或降雪的可能，则强烈建议不要依赖于来自摄像机或摄像机近旁位置的光照。许多光都可能被雨水和雪花直接反射回摄像机，使得分析工具无法发挥其功能。而如果有环境光的帮助，即使在较差的天气下，分析工具也有较高几率提供一些结果。

8.2 侦测距离

要确定基于AI的分析应用程序的最大侦测距离，是困难的 – 数据表中（以米或英尺为单位）给出的明确值无法代表全部事实。图像质量、场景特征、天气条件和目标属性（如颜色和亮度）都对侦测距离有着显著影响。举例来说，很明显的是，相较于雨天下的阴暗目标，晴朗天气下有着阴暗背景的明亮目标的视觉侦测距离要远得多。

侦测距离还取决于被侦测的目标的速度。如要获得准确的结果，视频分析应用程序需要在足够长的时间内“看到”目标。而这个时间的长短需要取决于平台的处理性能（帧速）：处理性能越低，侦测目标所需的这个可视时间越长。如果摄像机的快门时间与目标速度匹配不够良好，图像中出现的运动模糊也可能降低侦测准确度。

快速运动的目标在从摄像机近旁通过时可能更容易被丢失。例如，在摄像机远处奔跑的人可能被很好地侦测到，而以相同速度在距离摄像机近的地方奔跑的人可能快速进出摄像机视野，从而导致不会触发报警。

在基于运动侦测的分析中，直接朝向或远离摄像机运动的目标还带来了另一项挑战。对慢速运动目标的侦测将尤其困难，相较于跨整个场景的运动，这种情况产生的图像变化将会很小。

分辨率较高的摄像机通常不支持较长的侦测距离。执行机器学习算法所需的处理能力与输入数据的大小成正比。这就意味着，相比1080p摄像机，分析4K摄像机全分辨率所需的处理能力要至少高出四倍。常见的是，由于摄像机的处理能力限制，基于AI的应用程序在运行时所使用的分辨率要低于摄像机或数据流能够提供的分辨率。

8.3 报警与录像设置

由于应用了多种过滤级别，目标分析工具产生的假报警少。但只有在满足所列出的全部前提条件时，目标分析工具才能正确发挥其性能。否则，它们可能错失重要事件。

如果并不确定是否始终满足所有条件，那么建议采取保守措施，并将系统设置为：特定目标分类不作为仅报警触发因素。这将导致更多的假报警，但同时也会降低错失重要信息的风险。如果让报警或触发直接到达报警监控中心，假报警会造成高昂的代价。因此，明显需要一套可靠的目标分类系统来过滤掉不期望的报警。但录像解决方案仍然可以并且应该被设置成不仅仅依靠目标分类。在错失真报警的情况下，这种设置让您能够根据录像内容评估错失报警的原因，然后改进整个系统和配置。

如果在事件搜索期间，在服务器上执行目标分类，则建议将系统配置成能够持续录像，且根本不过滤原始录像内容。持续录像会占用大量存储空间，但这在一定程度上可通过先进的压缩算法（如Zipstream）予以弥补。

8.4 维护

应定期维护监控系统。建议开展实地检查，而不仅是通过VMS接口查看视频，以便发现并移除可能干扰或阻挡摄像机视野的物体。这在仅支持录像功能的标准系统中也是重要的，而在使用分析工具时，甚至更为重要。

在基础视频移动侦测中，典型的障碍物，比如随风摇摆的蜘蛛网，可能使报警数量增加，造成不必要地占用存储空间。在支持目标分析的情况下，蜘蛛网基本上会在侦测区域中形成一个排除区。它的威胁在于，有可能使目标变得模糊，大大降低侦测和分类的成功几率。



蜘蛛网可能干扰监控摄像机的视野。

摄像机前玻璃罩或圆罩上的污渍不大可能在白天引起问题。但在低照度条件下，从侧部投射到脏污圆罩上的光（比如，来自汽车前大灯）可能造成不期望的反射，进而可能降低侦测准确度。

场景相关的维护跟摄像机维护一样重要。在摄像机的寿命期内，监控场景内可能发生各种各样的情况。简单的前后图像比较将能够揭示潜在问题。部署摄像机时的场景状况如何，而如今的场景状况又是如何？是否需要调整侦测区域？是否应调整摄像机视野，或者是否应将摄像机转移到别的位置？

9 隐私与个人信息保护

安防与监控业务要求平衡隐私和个人信息保护方面的权利，通过防止犯罪或助推司法证据调查的方式，增强安全性。在特定的系统和应用场合中，这要求谨慎的道德伦理考量，以及了解并遵守当地法律法规。它还对解决方案提出了要求，例如，确保网络安全以及防止视频材料的意外泄露。同时，如果仅传输匿名数据以供后期分析，那么出于统计目的的前端分析和元数据生成便有助于提升隐私保护。

随着自动化分析在监控系统中的应用日益广泛，还必须考虑某些新的方面。由于分析应用程序伴随有假侦测风险，让有经验的操作人员或用户参与决策过程是很重要的。这通常被称为“人机回圈”。此外，还必须认识到，人类决策可能受到报警生成和呈现方式的影响。在未经恰当培训且未充分了解分析解决方案的功能的情况下，可能得出错误的结论。

另一关注点可能是深度学习算法的开发方式，对于某些应用场合，要求以谨慎的方式应用这一技术。这些算法的质量基本上与用于算法训练的数据集（即，视频和图像）相关。测试结果显示，如未仔细挑选材料，某些AI系统可能在侦测中同时表现出种族偏见和性别偏见。这引发了公开的讨论，同时，不仅在法律层面予以了相关限制，而且还要求采取措施确保在系统开发期间解决这样的问题。

随着AI在监控领域的应用日益广泛，严谨评估这项技术的使用场合与时机，从而在运行效率优势与新应用场合之间合理平衡，这是很重要的。

10 附录

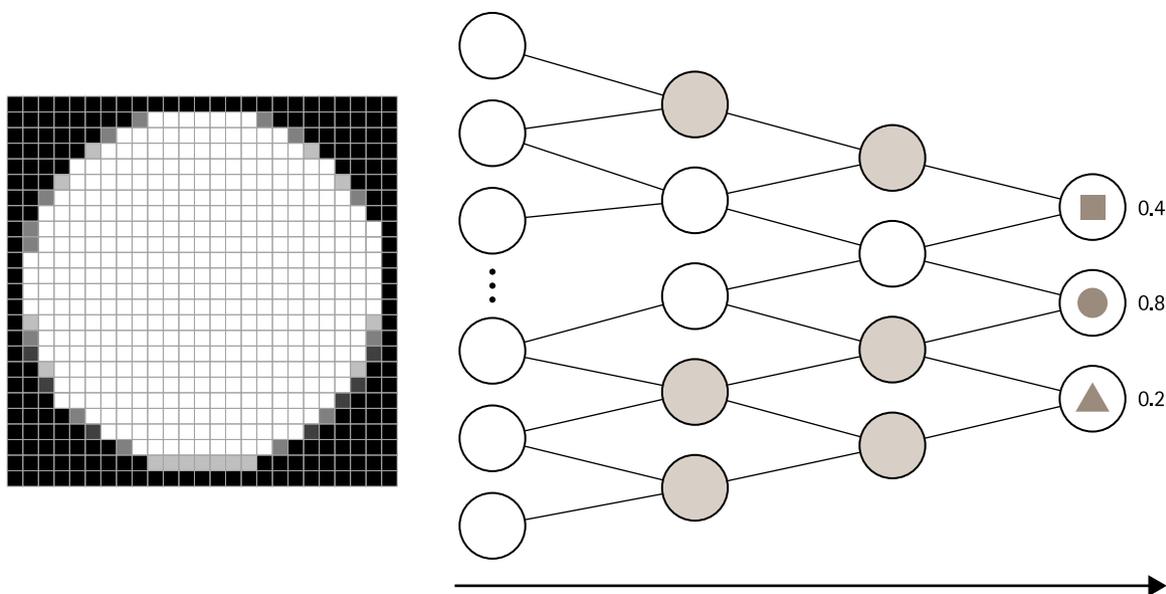
本附录介绍有关人工神经网络的背景信息，这些网络构成深度学习的基础。

10.1 神经网络

神经网络是一系列算法，这些算法用于通过在某种程度上与人脑工作方式类似的过程，识别数据集中的关系。神经网络是由多层所谓的节点或神经元互连形成的层级结构，信息通过网络沿各连接从输入层传送到输出层。

神经网络工作的假设条件是，能够将输入数据样本减少至有限特征集，从而更好地表征输入数据。这些特征随后可以被组合，并将有助于对（例如）描述图像内容的输入数据分类。

下图所示的示例使用了神经网络来识别输入图像所属的类别。图像中的每个像素均由一个输入节点来表征。所有输入节点都连接到第一层的节点。由此生成的输出值被作为输入值一路传送至第二层，以此类推。在每一层中，过程还涉及加权函数、偏差值和激活函数。



输入图像（左）和神经网络（右）的示例。当到达输出层时，网络已经推断了每种可能类别（正方形、圆形或三角形）的概率。拥有最高概率值的类别是输入图像的更可能的形状。

这个过程被称为**前向传播**。如果前向传播的结果不匹配，则会通过**反向传播**略微修改网络参数。在这个迭代训练过程中，网络性能逐渐提升。

在部署之后，神经网络通常没有对先前前向传输的记忆。这就意味着，在一段时间后，它将不会改进，并且它将仅能够侦测所训练的目标类型，或者解决所训练的任务类型。

10.2 卷积神经网络 (CNN)

卷积神经网络 (CNN) 是人工神经网络的一种子类型，经验证尤其适用于计算机视觉任务，它们是快速深度学习的核心所在。就计算机视觉而言，网络被训练成能够自动寻找有特色的图像特征，比如边缘、拐角和色差，以便在整个图像上有效识别目标形状。

实现这一目的的主要操作是一种被称为卷积的数学运算。这是一种非常高效的运算，因为每个节点的输出仅取决于输入数据中的有限环绕，这种有限环绕由前一层形成，而不是使用全部输入数据来形成。换言之：在CNN中，每个节点并不连接到前一层中的每个节点，而是仅连接到小的子集。卷积可辅以能够降低数据大小并同时保留有用信息的其他运算。与在标准神经网络中一样，数据进入网络越深，就变得越抽象。

在训练阶段，CNN学习层应用的更佳方式。也就是说，卷积应如何组合来自前一层的特征，以便使网络输出尽可能与训练数据的注释一致。然后，在推断期间，经训练的卷积神经网络依次应用作为训练结果的卷积层。

关于安讯士(Axis Communications)

安讯士通过打造网络解决方案，不断提供改善安防技术的独特见解并引入创新业务模式，旨在创造一个更加智能、安全的世界。作为网络视频行业的领导者，安讯士致力于推出视频监控和分析应用、访问控制、内通系统以及音频系统的相关产品和服务。安讯士在全球50多个国家和地区设有办事机构，拥有超过3,800名尽职的员工，并与遍布世界各地的合作伙伴携手并进，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。

关于安讯士的更多信息，请访问我们的网站：axis.com