Surveillance video/image data collection and analysis based on AI and IOT technologies has been an important technical grip for different scenes in smart city development. The trust information extraction in surveillance data has aroused many researchers' interest and motivates the analysis of images from numerous IoT visual sensors [1-3]. However, the massive deployment of visual sensors leads to a number of challenges: firstly, the sheer volume of camera video images leads to a data disaster. With 30 frames per second and 5mb per image, a single camera generates a data storage requirement of 12,656.25Gb in a day, while IHS research indicates that there will be over 1 billion surveillance cameras worldwide after 2021. These video stores take up a lot of hardware resources and no data centre can afford the daily growth of video data and has to be covered periodically [4]. Secondly the redundancy of information in massive camera video data leads to key information overwriting and difficulties in retrieving information based on video []. In addition, massive video transmission occupies a large amount of communication bandwidth, with high communication costs, making it difficult to achieve collaborative use of a large range of cameras for mega-city governance. With the development of AI technology, surveillance cameras have led to the leakage of residents' biometric privacy, raising ethical and regulatory concerns. How to safeguard the function of surveillance cameras while ameliorating the above challenges has become a direction of research for a wide range of scholars.

In this paper, we take a human-like cognitive perspective to carry out based theory research for exploring new models of large-scale camera urban applications. We humans, from infants to the elderly, perceive a large amount of picture information with both eyes over decades and can have long-term clear memories of the people and things we experience and when and where they happened. However, we are often unable to reproduce all of the image information that occurred, but rather combine it with high-dimensional semantic abstraction to achieve coarse-grained picture recall. Our memories of familiar faces of friends or family members are also often not remembered through images of faces, nor are they even identified by specific features such as single eyelids and double eyelids, but rather translated into a general impression of high-dimensional semantic information. In addition, the high-dimensional abstract semantics in our human brain memory plays an important role in blurring human decryption recognition, where humans can recognise the identity of people we are familiar with in blurred or partially obscured images of faces, whereas strangers have difficulty in doing so. This is a result of the long biological evolution of humans to cope with the challenges of massive amounts of AV data, and it is difficult to have a theoretical explanation of the mechanism by which human low-dimensional fine-grained information is correlated to higher-dimensional coarse-grained information. There are even fewer attempts to use the mechanism to implement video processing for massive surveillance cameras. This paper presents a preliminary exploration in this direction, proposing the study of face degradation encryption in video images and then combining high-dimensional semantic information with identity recognition for autonomous decryption, and achieving high approximation face recovery.

随着经济全球化的影响和城市化进程的加快，城市人口密度上升，人流和车流增加，城市建筑和基础设施的布局越来越复杂。这就导致了城市建设中的交通、社会治安等城市管理问题。近年来，治安形势的多样性和复杂性逐渐增强，犯罪手段越来越隐蔽，这对城市的治安管理提出了新的挑战。在治安管理中，分布在城市各个角落的监控摄像头发挥着不可或缺的作用。基于人工智能和物联网技术的监控视频/图像数据采集和分析已经成为智慧城市发展中不同场景的重要技术抓手。

监控数据中的信任信息提取引起了许多研究者的兴趣，并促使人们对来自众多物联网视觉传感器的图像进行分析[1-3]。然而，视觉传感器的大规模部署导致了一系列的挑战：1）首先，摄像机视频图像的巨大数量导致了数据灾难。以每秒30帧、每张图片5MB计算，一个摄像头一天产生的数据存储需求为12656.25Gb，而IHS的研究表明，2021年后全球将有超过10亿个监控摄像头。这些视频存储占用了大量的硬件资源，没有一个数据中心能够承受每天增长的视频数据，必须定期进行覆盖[4]。2）其次海量摄像机视频数据中的信息冗余导致关键信息被覆盖，基于视频的信息检索困难[]。3）此外，海量视频传输占用了大量的通信带宽，通信成本高，难以实现大范围摄像头的协同使用，难以实现超大型城市治理。[]4）同时监控摄像头导致了居民生物识别隐私的泄露，引发了伦理和监管方面的担忧，因此遮蔽或掩盖人脸技术成为视频监控的迫切需求。如何保障监控摄像头的功能，同时改善上述挑战，成为广大学者的研究方向。

在本文中，我们从类似人类的认知角度出发，开展基于理论的研究，探索大规模摄像头城市应用的新模式。我们人类，从婴儿到老人，几十年来用双眼感知了大量的图片信息，并能对所经历的人和事以及发生的时间和地点有长期清晰的记忆。然而，我们往往无法再现所有发生的图像信息，而是将其与高维语义抽象相结合，以实现粗粒度的图片记忆。我们对朋友或家人的熟悉面孔的记忆也往往不是通过面孔的图像来记忆的，甚至也不是通过单眼皮和双眼皮等具体特征来识别，而是转化为高维语义信息的一般印象。此外，我们人类大脑记忆中的高维抽象语义在模糊的人类解密识别中起着重要作用，人类可以在模糊或部分模糊的人脸图像中识别出我们熟悉的人的身份，而陌生人则很难做到这一点。这是人类长期生物进化的结果，以应对海量AV数据的挑战，而对于人类低维细粒度信息与高维粗粒度信息的关联机制，很难有理论上的解释。利用该机制来实现大规模监控摄像机的视频处理的尝试就更少了。本文在这个方向上做了初步探索，提出研究视频图像中的人脸降级加密，然后将高维语义信息与身份识别相结合进行自主解密，并实现高近似度的人脸恢复。