**With the large-scale industry application of artificial intelligence and video surveillance, massive video data storage and personal privacy issues are highlighted, which restrict the application expansion. From the perspective of humanoid memory mechanism, we propose a video abstraction encryption and decryption algorithm based on high and low dimensional information association cognitive mechanism, which uses face recognition algorithm to locate human faces and encrypt the original video with mosaic encryption method, and then perform spatio-temporal index encoding, and further use abstract face feature memory to match and decode with the same identity person to construct a de-mosaic decryption key function. The core innovation is the humanoid memory mechanism for parsing and modelling, and combined with specific AI techniques such as YOLO and GAN for initial experimental validation in face encryption and decryption. The algorithm research will be important in the research of video information compression and storage, person re-identification and personal privacy protection.**

The surveillance cameras distributed in all corners of the city play an indispensable role in the city security management. 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 extraction of trust information from surveillance data has attracted the interest of many researchers and has led to the analysis of images from numerous IoT vision sensors [1-3]. However, the large-scale deployment of vision sensors leads to a number of challenges: 1) First, the huge number of camera video images leads to a data disaster. At 30 frames per second and 5MB per image, a single camera generates a data storage requirement of 12,656.25Gb a day, while IHS research indicates that there will be over one billion surveillance cameras worldwide in future. These video stores take up a large amount of hardware resources, and no data center can withstand the daily growth of video data, which must be overwritten on a regular basis [2]. Secondly, information redundancy in massive camera video data leads to key information being overwritten and video-based information retrieval being difficult [3]. In addition, massive video transmission takes up a large amount of communication bandwidth, and communication costs are high, making it difficult to achieve widespread cameras for collaborative use to achieve mega-city governance [4]. Meanwhile surveillance cameras have led to the leakage of residents' biometric privacy, raising ethical and regulatory concerns. How to safeguard the functionality of surveillance cameras while improving the above challenges has become a research direction for a wide range of scholars.

In this paper, we take a humanoid cognitive perspective to carry out theoretical 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 experienced. 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. We also tend to remember familiar faces not through detailed facial features such as single or double eyelids, but into general impressions of higher-dimensional semantic information. In addition, the high-dimensional abstract semantics in our human brain memory plays an important role in blurring human decryption recognition. Humans can recognize acquaintances through blurred or partially blurred facial images, but not strangers. The process of humanoid perceptual memory mechanism to handle the massive amounts of video data is difficult to have a theoretical explanation. But the association between low-dimensional fine-grained information and higher-dimensional coarse-grained information for humanoid perceptual data compression and decryption has theoretical significance and practical value,

which are worth using for processing massive surveillance vedio data. In this paper, we try to propose an autonomous face degradation encryption and decryption algorithm based on the above humanoid association memory mechanism.

In the perspective of recent advances in the field of AI-driven face recognition of video surveillance, the human face object tracing for video surveillance has gained widespread adoption in urban security and community management. A lot of scholars are committed to the research of computer vision technique with promising accuracies and efficiencies for face recognition and object detection [5-7]. The face recognition methods mainly include 1) traditional methods, which rely on hand-crafted feature extraction techniques and a pre-trained classifier along with fusion, and 2) deep learning methods, which automatically learn features and classifiers together utilizing enormous quantities of data[10,13, 14]. With the development of deep learning technology, the application boundary of face recognition will be gradually opened. The majority of face recognition in video surveillance today is "closed-set," which only recognizes the identity of previously registered objects. However, "open-set" has gained popularity as a result of the differences between the source and target domains, which make it less effective when transferring face recognition systems from controlled environments to uncontrolled scenes. Suandi proposed fuzzy ARTMAP neural networks to solve the open-set single-sample face recognition problem and an automatic pose normalization technique without model fitting and human intervention, which greatly improves the performance of open-set single-sample face recognition methods in surveillance environments [9,11]. The "open-set" face recognition prone to increase the human privacy exposure degree in the ubiquitous city surveillance network.

The low resolution of urban monitoring picture and the difficulty of small face feature extraction are being changed. Even though the surveillance cameras are usually placed far away from the objects and the resolution of the captured face images is low due to distance, extensive research has been carried out for recognizing acceptable recognition features at low quality video frames. Zhao et al. took an end-to-end approach to match high-resolution (HR) images with low-resolution (LR) images in surveillance videos[8]. Singh et al. improved the number of descriptors in the image and mitigates the effects of noise based on super-resolution faces[12]. Dharrao et al. used the Viola-Jones algorithm to detect the face part in the video sequential frames and improved the quality of the face part by applying a super-resolution scheme based on bicubic interpolation[15]. In addition, the multi-resolution convolutional neural networks (MRCNN) and anti-aliasing techniques were adopted to solve the low-resolution problems[16].

The development trend of face recognition technologies shows that the challenge of citizen's face privacy feature under the ubiquitous cameras is more and more serious. How to explore a new paradigm for large-scale camera urban applications from the perspective of humanoid cognition by performing face reduction encryption on the recognized video images are meaningful.

## A. Problem description.

The process of human face perception and identity recognition based on fuzzy impression memory association is highly complex. Each of us sees many faces in daily life scenarios, however, not all the information about faces are remembered. As shown in figure 1 for example, when some people meet with each other unintentionally, their mind will unconsciously recall that they have seen such a face at a certain time, place and event. Moreover, they can recall the memory of more detailed scene and clearer features. The process can actually be simplified as the human brain perceives the concrete face image information seen by the eyes to extract high-dimensional abstract semantic features. The high-dimensional abstract semantic features are retrieved and matched with the high-dimensional semantic information indexed in memory combining person, event, time and place, and the past feature-blurred memory scene is clearly reproduced in combination with the current perceived face image.

The general expression of the Humanoid Association is as follows.

(1)

I是图像

Ai 是大脑在第一阶段感知到的第 i 张人脸的低维全维度信息集

Bi 是大脑在 Ai 的基础上形成的第 i 张人脸的高维抽象语义特征集、

Aip是Ai的加密集

A'ip是由A\_ip和B\_i部分组成的解密集，

A'〗\_i是大脑在第二阶段感知到的关于人脸的低维全维度信息集，

B'〗\_i是大脑在头脑中根据〖A'〗\_i形成的高维抽象语义特征集。

保留高斯模糊的部分

将提取密钥的部分和所谓的加解密的部分分开

在一个真实的现实场景下，只对人脸进行加解密，也就是说我们有出现过的全部人脸的密钥池（云端），

有一段视频监控，将人脸的部分定位出来，加密以后会有一段人脸的部分是模糊的视频，也就是加密以后的视频；这段加密以后的视频是存储在本地的

1. 有一个年轻时候的人脸，先经过高斯加密，得到一个模糊的人脸，存储在本地
2. 将这个年轻的人脸，上传到云端通过psp的encoder抽取密钥，形成人脸密钥特征池，存储在云端
3. 当需要调取某一个人的监控，获得了他当前的人脸照片，相当于就是年老以后的人脸，
4. 对当前的年老以后的人脸同样经过psp的encoder抽取密钥
5. 拿着这个密钥去跟云端密钥池中的密钥进行匹配，找到了匹配的密钥
6. 这个密钥附带有时空索引等，所以只要知道是哪个密钥就知道要解密的部分是哪里，然后将这个密钥放到了styleGAN的生成器中生成解密以后的人脸，并用这个人脸将视频对应区域进行还原，就可以得到解密之后的视频了。
7. 本地存储的是高斯模糊以后的视频，这个视频是没有办法直接解密的，所以必须通过GAN来生成解密之后的人脸，从而才能够进行人脸的恢复.

那为什么我不直接对加密以后的视频进行全部解密呢，如果我加密以后的视频只是进行了一个高斯模糊的话，但是不可以，这里没有办法解密因为我们要保护其他人的隐私，所以我们要针对性地只解密这个人的人脸，那我们怎么知道到底哪个是这个人的人脸的区域呢？我们对当前的年老以后的人脸同样经过psp的encoder抽取密钥，拿着这个密钥去跟云端密钥池中的密钥进行匹配，找到了匹配的密钥，相当于就知道这个人是谁了，

这个密钥附带有时空索引等，所以只要知道是哪个密钥就知道要解密的部分是哪里，然后将这个密钥放到了styleGAN的生成器中生成解密以后的人脸，并用这个人脸将视频对应区域进行还原，就可以得到解密之后的视频了。

本地存储的是高斯模糊以后的视频，这个视频是没有办法直接解密的，所以必须通过GAN来生成解密之后的人脸，从而才能够进行人脸的恢复.

与传统的视频监控系统不同，这种方法不再存储原始视频，而是选择本地存储加密视频，同时将高维抽象语义上传到云端进行后续处理和分析。例如，对于监控后的人脸检索服务，由于本地存储加密后的人脸丢失了大量的人脸特征信息，因此无法对视频进行检索复核，而需要对高维语义B进行索引以便查询服务，这与人类的感知记忆类似。

为了使人工智能系统具备类似人类的高维抽象计算能力，我们利用（1）中的模型来提取高维语义。中的模型从原始视频中提取高维抽象语义特征 B，并将视频帧的时序、边缘摄像头本身的经纬度和像素坐标融合为高维抽象语义特征的符号位编码，并将高维抽象语义特征 B 与加密的人脸图像帧之间的映射关系与上述符号位编码关联起来，以方便后续的视频解密工作。在训练过程中，提取视频人脸数据集的抽象语义特征 B1 作为解码密钥 A1p，并存储在云密钥库中。

上述模型学习到的高维抽象语义特征往往无法解释，因此需要对抽象语义 B1 和 B2 进行深入研究，研究如何通过抽象语义 B1 和 B2 进行人脸匹配检索。目前，在人脸识别领域，通过深度神经网络匹配人脸特征来判断人脸身份的技术已经相对成熟，而利用抽象语义 B1 和 B2 作为身份关键来判断人脸身份则变得更具挑战性。

人脸的身份识别分为两部分，一部分是通过匹配人脸的具体特征来确定身份，而不是简单地通过比较具体特征来确定身份；另一部分属于印象抽象语义，就像人脸给人的总体印象一样，通过匹配人脸的具体特征来确定身份。当需要对人脸进行解密时，首先使用相同的模型提取人脸的特征，然后采取一定的维度与人云密钥库中的特征进行匹配，并计算匹配的准确率。这样就实现了召回触发调用函数 f2。

我们使用 pixel2style2pixel（pSp）模型对 YOLO5-face 定位的人脸进行加密和编码。加密人脸以获得高维抽象语义 B1 的模型表达式如下：

B\_1=L(E(A\_1 )+w ̅)

以之前的原始视频 A1 为输入，L(\*) 表示获取 A1 的潜码，从而得到抽象语义特征 B1；E(\*) 表示 pSp 模型的编码器，由 E(A\_1 ) 得到的势向量与网络模型中的平均势向量 w ̅ 相加，得到最终的势向量。这一步骤通常有助于平衡生成图像的质量和多样性。此外，抽象语义特征 B1 被用作解码 A1p 的密钥，并存储在云密钥库中。至此，f1(Ai) 的求解过程就完成了。

After completing face target detection, the faces in the recognized frames are encrypted. We use the pixel2style2pixel (pSp) model to encrypt and encode the faces localized by YOLO5-face.The model expression for encrypting the face to obtain the high dimensional abstract semantics B1 is as follows:

)

Taking the previous original video A1 as input, L(\*) denotes the latent code of obtaining A1 to get the abstract semantic feature B1.E(\*) denotes the Encoder of the pSp model.The potential vector obtained from E(A\_1 ) is summed with the average potential vector w ̅ in the network model to obtain the final potential vector. This step usually helps to balance the quality and diversity of the generated images. Further, the abstract semantic feature B1 is used as the key to decode A1p and stored in the cloud keystore. The solving process of f1(Ai) is thus completed.

A. 视频监控的人脸识别

从人工智能驱动的视频监控人脸识别领域的最新进展来看，用于视频监控的人脸目标追踪已在城市安防和社区管理中得到广泛应用。很多学者都致力于研究计算机视觉技术在人脸识别和物体检测方面具有良好的精度和效率[5-7]。人脸识别方法主要包括：1）传统方法，即依靠手工特征提取技术和预先训练的分类器进行融合；2）深度学习方法，即利用海量数据自动学习特征和分类器[10,13,14]。随着深度学习技术的发展，人脸识别的应用边界将逐渐被打开。目前，视频监控中的人脸识别大多是 "封闭集"，只能识别先前注册对象的身份。然而，由于源域和目标域之间存在差异，当人脸识别系统从受控环境转移到非受控场景时，其效果会大打折扣，因此 "开放集 "逐渐受到青睐。Suandi提出了模糊ARTMAP神经网络来解决开集单样本人脸识别问题，并提出了一种无需模型拟合和人工干预的自动姿态归一化技术，大大提高了开集单样本人脸识别方法在监控环境下的性能[9,11]。在无处不在的城市监控网络中，"开放集 "人脸识别容易增加人的隐私暴露程度。

城市监控图像分辨率低，小范围人脸特征提取困难，这些问题正在得到改变。尽管监控摄像头通常距离拍摄对象较远，拍摄到的人脸图像分辨率也因距离而较低，但人们仍对如何在低质量视频帧中识别可接受的识别特征进行了大量研究。Zhao 等人采用端到端方法匹配监控视频中的高分辨率（HR）图像和低分辨率（LR）图像[8]。Singh 等人改进了图像中描述符的数量，并基于超分辨率人脸减轻了噪声的影响[12]。 Dharrao 等人使用 Viola-Jones 算法检测视频连续帧中的人脸部分，并通过应用基于双三次插值的超分辨率方案提高了人脸部分的质量[15]。此外，还采用了多分辨率卷积神经网络（MRCNN）和抗锯齿技术来解决低分辨率问题[16]。城市监控图像分辨率低，小范围人脸特征提取困难，这些问题正在得到改变。尽管监控摄像头通常距离拍摄对象较远，拍摄到的人脸图像分辨率也因距离而较低，但人们仍对如何在低质量视频帧中识别可接受的识别特征进行了大量研究。Zhao 等人采用端到端方法匹配监控视频中的高分辨率（HR）图像和低分辨率（LR）图像[8]。Singh 等人改进了图像中描述符的数量，并基于超分辨率人脸减轻了噪声的影响[12]。 Dharrao 等人使用 Viola-Jones 算法检测视频连续帧中的人脸部分，并通过应用基于双三次插值的超分辨率方案提高了人脸部分的质量[15]。此外，还采用了多分辨率卷积神经网络（MRCNN）和抗锯齿技术来解决低分辨率问题[16]。

隐私泄露问题已引起广泛关注。视频监控的人脸识别在日常生活中已无处不在，但智能视觉应用与个人隐私保护之间却难以平衡。除了完善相关法律法规来规范视频的采集、存储和使用外，还需要相应的技术措施来保护个人隐私。基于密码学的人脸隐私保护方案对视频中显示身份的人脸区域进行选择性加密，在未来有合法需求时可解密恢复原始视频。如何将仿人联想记忆机制的自主人脸降级加解密算法融入人工智能人脸识别算法，是一个亟待突破的方向。

Most of the existing face encryption schemes are homomorphic-based[18-26].