A Face Privacy Protection and Self-decryption Method Based on Humanoid Association Mechanism

摘要：  
随着人工智能和视频监控的大规模行业应用，海量视频数据存储和个人隐私问题凸显，制约了应用拓展。我们从类人记忆机制的角度出发，提出了一种基于高低维信息关联认知机制的视频抽象加解密算法，利用人脸识别算法定位人脸，采用马赛克加密方法对原始视频进行加密，然后进行时空索引编码，并进一步利用抽象人脸特征记忆与同一身份人进行匹配解码，构建去马赛克解密密钥函数。其核心创新点是采用类人记忆机制进行解析和建模，并结合pixel2style2pixel模型提取某几维特征与人脸特征密钥池中的特征去匹配，计算匹配的准确度，在人脸加解密方面进行了初步实验验证。该算法研究将对视频信息压缩与存储、人物再识别和个人隐私保护等方面的研究具有重要意义。

**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 human-like 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.**

# INTRODUCTION

基于人工智能和物联网技术的监控视频/图像数据收集与分析一直是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人员的兴趣，并促使他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带来了一系列挑战：首先，摄像头视频图像的海量导致了数据灾难。以每秒 30 帧、每幅图像 5mb 的速度计算，单个摄像头每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，任何数据中心都无法承受每天视频数据的增长，必须定期覆盖[4]。其次，海量摄像头视频数据中的冗余信息导致关键信息被覆盖，基于视频的信息检索困难重重[]。此外，海量视频传输占用大量通信带宽，通信成本高，难以实现超大城市治理中大范围摄像头的协同使用。随着人工智能技术的发展，监控摄像头导致居民生物特征隐私泄露，引发伦理和监管方面的担忧。如何在保障监控摄像头功能的同时改善上述挑战，已成为广大学者的研究方向。

(类人认知角度)

本文从类人认知角度出发，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，都能在几十年的时间里用双眼感知大量的图像信息，并能对所经历的人和事及其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过的图像信息，而是将其与高维语义抽象结合起来，实现粗粒度的图像记忆。我们对朋友或家人熟悉面孔的记忆，往往也不是通过面孔的图像，甚至也不是通过单眼皮、双眼皮等具体特征来识别，而是转化为高维语义信息的总体印象。此外，人脑记忆中的高维抽象语义在模糊人脸解密识别中也发挥了重要作用，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。这是人类长期应对海量视听数据挑战的生物进化结果，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解释。利用该机制实现海量监控摄像机视频处理的尝试更是少之又少。本文在这个方向上进行了初步探索，提出研究视频图像中的人脸降维加密，然后结合高维语义信息和身份识别进行自主解密，实现高近似度的人脸恢复。

# RELATED WORK.

1. Encryption method

# III. PROPOSED APPROACH

## We describe the human brain face perception-memory-association re-identification process through a typical scenario and outline the scientific problem to be solved in this paper. An algorithmic model for solving the problem and a framework for the solution process are then proposed.

我们通过一个典型场景描述了人脑面部感知-记忆-联想重识别过程，并概述了本文要解决的科学问题。 然后提出了解决问题的算法模型和解决过程的框架。

A、问题描述。

人类根据模糊印象记忆关联感知面孔并识别其身份的过程非常复杂。 我们每个人在生活场景中都会看到很多面孔，比如在拥挤的人群中、在参加的学术会议上、在我们见面的晚宴上等等。然而，并不是所有关于面孔的信息都被普遍记住。 然而，当面对面结识新朋友时，人们倾向于在大脑会话中寻找似曾相识的印象。 让我们从一个典型场景开始描述这个过程。 如图1所示，我们在学术会议上遇到了很多人，但并不是每个人都互相交流。 当将来的某个时候，你们两个再见面的时候，你的脑海里会不自觉地想起，你在某个时间、某个地点、某件事上见过这样的一张脸，再结合眼前的真人，你似乎就想起了那个人。 场景并更清楚地了解他的特征。 这个过程其实可以简化为人脑感知眼睛看到的具体人脸图像信息，提取高维抽象语义特征，检索高维抽象语义特征并与内存中索引的高维语义信息进行匹配 将人物、事件、时间、地点结合起来，结合当前感知到的人脸图像，清晰地再现过去特征模糊的记忆场景。

模型的一般表达式如下。

其中，A1 是大脑当前感知到的关于人脸的低维全维度信息集合，B1 是大脑根据 A1 在头脑中形成的高维抽象语义特征集合，A2p 是大脑中模糊记忆的部分低维残余信息集合，A2 是大脑当前感知到的关于人脸的低维全维度信息集合，A2p 是 A2 的子集，B2 是大脑根据 A2 在头脑中形成的高维抽象语义特征集合。求解上述表达式的算法如下。

之前获得了A1和B1，在此基础上，当再次获得A2p和B2的情况下，可以解码得到A2

输入：A1, A2p, B2

输出：A2

然后，有一些算法的步骤被描述如下：

对于A1在大脑中做如下：

使用脑神经网络进行高维语义抽象

f1(Ai) → AiP, Bi，其中 A1P 是 A1 的一个子集，i=1, 2, ……

对于B1在大脑中做如下：

在B1中寻找B2

找到最大脸部相似性特征

P(A2p| A1p)= f2(B2 ∩ B1 )→1

对于A2p在大脑中做如下：

f3(A1, A2p, B2) → A2

返回 A2

本文将上述类人感知联想记忆算法与监控视频的人脸加解密要求相结合，解决了以下问题。

1) 借鉴类人认知机制，建模求解高维抽象记忆和压缩感知过程 f1 函数，提出求解 Aip 和 Bi 的人工智能算法，以识别定位视频中的人脸，提取高维语义特征，同时对分辨率降低的视频人脸进行加密。

2）借鉴类人联想记忆机制，对高维语义特征的记忆存储和联想匹配 f2 进行建模，提出基于 B1 和 B2 的召回触发匹配索引机制，实现联想记忆匹配。

3) 借鉴类人感知触发式调用机制，模拟高维语义特征 A1 和 B2 低分辨率视频的联想调用，解决索引位置以及高分辨率解密的 f3。

A1认为是人脸原图

F1(Ai)->不太清晰的模糊的人脸Aip；高维抽象之后的特征Bi

F1函数相当于人脸加密编码函数，得到了存储在密钥

当你再次看到一个人时，我可以认为你有一个新的人脸原图的输入A2,

通过一样的脑神经网络你可以得到一样经过一定抽象的A2p和B2

F2召回触发式调用机制，f2其实是匹配函数的实现，如何做特征匹配的

对新得到的B2和之前人脑中的存储的B1集合进行特征点匹配，

证明说使用越高维的特征匹配效果越好，那么说明我们的猜想的可实现性

F3相当于解密函数，就是匹配找到密钥以后，通过密钥去做人脸还原，得到还原以后的人脸图像

## B. AI Methodologies

受到类人感知、压缩记忆和联想记忆的启发，我们提出了一种算法框架，可用于加密/解密监控视频中的人脸，如下图所示。

**1) Encryption method**

对于视频帧输入V，使用YOLO5-face深度学习模型ф实现边缘端监控摄像头对人脸的识别和定位，得到A1=ф（V），选择YOLO5-face是因为该模型针对人脸识别分割需求，在YOLOV5中增加了地标分支，通过五个人脸关键点回归翼损失函数，提高了人脸检测和定位的准确性。

完成人脸目标检测后，对识别帧中的人脸进行高斯模糊处理，即对人脸图像进行高斯卷积预算，其概率密度分布函数如下图所示。

问题是我们所谓的高斯模糊其实是通过psp模型去进行的加密，也就是没有所谓的高斯模糊.

这里是我修改以后的部分：

完成人脸目标检测后,对识别帧中的人脸进行加密并存储得到的密钥。与传统视频监控系统不同，该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，我们使用 pixel2style2pixel (pSp)模型的Encoder,用于后续分析 YOLO5-face 定位的人脸，提取高维抽象语义特征 B，并融合视频帧时间序列、 将边缘摄像头自身的经纬度和像素坐标作为高维抽象语义特征的标志位编码导入云数据库，高维抽象语义特征 B 与加密后的人脸图像帧之间的映射关系可与上述标志位编码关联，便于后续视频解密工作的开展。

获取高维抽象语义 B1 的模型表达式如下。

B1=L(E(A1)+W).

将之前的原始视频A1作为输入，L(\*)表示获得A1的latent code，得到抽象语义特征B1.E(\*)表示pSp模型的Encoder, 它在ResNet的基础上，使用标准的feature pyramid 来提取特征映射。

对于18个目标风格中的每一个， 都有一个小的mapping net 被训练来 从对应的特征映射中提取学习到的风格。 其中，0-2 风格 是从small 特征映射中生成的，3-6 风格是从 medium 特征映射中生成的， 7-18 是从largest 特征映射中生成的。借鉴类人认知机制，与人脑高维高维抽象记忆和压缩感知过程对应，从越小的特征映射生成的是越高维的特征信息。将E(A1)得到的潜在向量与网络模型中的平均潜在向量w相加，以得到最终的潜在向量。这一步通常有助于平衡生成的图像质量和多样性。进而将抽象语义特征 B1 作为解码 A1p 的密钥，并将其存储在云端密钥库中。这样，f1(Ai) 的解码过程就完成了。

2) 密钥存储与匹配

与传统视频监控系统不同，该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，使用1)中的模型对原始视频提取高维抽象语义特征 B，并融合视频帧时间序列、 将边缘摄像头自身的经纬度和像素坐标作为高维抽象语义特征的标志位编码导入云数据库，高维抽象语义特征 B 与加密后的人脸图像帧之间的映射关系可与上述标志位编码关联，便于后续视频解密工作的开展。

pixel2style2pixel (pSp). Our pSp framework is based on a novel encoder network that directly generates a series of style vectors which are fed into a pretrained StyleGAN generator, forming the extended W+ latent space.

获取高维抽象语义 B 的计算过程如下。通过之前的原始视频 A1 和加密视频 A1p 形成配对数据集，A1p 加上随机噪声 э 作为生成器的输入进行训练学习，而 A1 则作为判别器的输入进行判别，使得 A1p +э→A1 。在训练过程中，我们在每个生成网络层中嵌入一个具有正交基的子空间模型，用于获取训练模型的层次语义，进而将抽象语义特征 B1 作为解码 A1p 的密钥，并将其存储在云端密钥库中。这样，f1(Ai) 的解码过程就完成了。

把这里的编码过程替换掉（从理论的角度解释好就好）.

上述模型学习到的高维抽象语义特征往往不具备可解释性，因此要研究如何通过抽象语义 B1 和 B2 进行人脸检索匹配还需要对抽象语义 B 进行深入研究。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较为成熟，而将抽象语义 B1 和 B2 作为身份密钥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，一部分是通过人脸的具体特征进行比对，而不是简单地通过比对具体特征来确定身份；另一部分则属于印象抽象语义学，就像人脸给人的总体印象一样，通过人脸的具体特征进行比对来确定身份。在需要解密人脸时，首先用同样的模型对人脸提取特征后，取某几维跟人云端密钥库中的特征去匹配，计算匹配的准确度。

3) 解密过程

上一节实现了原始视频加密以及密钥存储和查询匹配过程。本节讨论如何通过加密视频 A1p、高维抽象语义 B1 和第二段视频中与第一段视频中人物身份匹配的人脸 A2 输出解密后的视频 A1。

这一过程相当于联想记忆，我们可以将过去的图像与当前的图像联系起来，人脸模糊的特征往往可以重新清晰起来。

为此，我们建立了一个开放集人脸再识别和基于 styleGAN生成器的解密模型。首先，通过上一节介绍的方法，从 A2 中提取高维抽象语义 B2。然后，计算 B2 与加密视频 A1p 对应的密钥池中所有高维抽象语义 B1i 的相似度，取相似度最高的 B1i。如果相似度低于某个阈值，则判定该人脸为陌生人脸，拒绝进一步解密；如果相似度超过某个阈值，则将高维抽象语义 B1i 作为约束条件添加到以 A1p 为输入的生成器中，并输出解密后的视频 A1。

解密的过程其实就是将找到的匹配的特征再喂给的styleGAN的生成器

class GradualStyleEncoder(Module):  
 def \_\_init\_\_(self, num\_layers, mode='ir', opts=None):  
 super(GradualStyleEncoder, self).\_\_init\_\_()  
 assert num\_layers in [50, 100, 152], 'num\_layers should be 50,100, or 152'  
 assert mode in ['ir', 'ir\_se'], 'mode should be ir or ir\_se'  
 blocks = get\_blocks(num\_layers)  
 if mode == 'ir':  
 unit\_module = bottleneck\_IR  
 elif mode == 'ir\_se':  
 unit\_module = bottleneck\_IR\_SE  
 self.input\_layer = Sequential(Conv2d(opts.input\_nc, 64, (3, 3), 1, 1, bias=False),  
 BatchNorm2d(64),  
 PReLU(64))  
 modules = []  
 for block in blocks:  
 for bottleneck in block:  
 modules.append(unit\_module(bottleneck.in\_channel,  
 bottleneck.depth,  
 bottleneck.stride))  
 self.body = Sequential(\*modules)  
 self.styles = nn.ModuleList()  
 self.style\_count = opts.n\_styles  
 self.coarse\_ind = 3  
 self.middle\_ind = 7  
 for i in range(self.style\_count):  
 if i < self.coarse\_ind:  
 style = GradualStyleBlock(512, 512, 16)  
 elif i < self.middle\_ind:  
 style = GradualStyleBlock(512, 512, 32)  
 else:  
 style = GradualStyleBlock(512, 512, 64)  
 self.styles.append(style)  
 self.latlayer1 = nn.Conv2d(256, 512, kernel\_size=1, stride=1, padding=0)  
 self.latlayer2 = nn.Conv2d(128, 512, kernel\_size=1, stride=1, padding=0)  
 def \_upsample\_add(self, x, y):  
 \_, \_, H, W = y.size()  
 return F.interpolate(x, size=(H, W), mode='bilinear', align\_corners=True) + y  
 def forward(self, x):  
 x = self.input\_layer(x)  
 latents = []  
 modulelist = list(self.body.\_modules.values())  
 for i, l in enumerate(modulelist):  
 x = l(x)  
 if i == 6:  
 c1 = x  
 elif i == 20:  
 c2 = x  
 elif i == 23:  
 c3 = x  
 for j in range(self.coarse\_ind):  
 latents.append(self.styles[j](c3))  
 p2 = self.\_upsample\_add(c3, self.latlayer1(c2))  
 for j in range(self.coarse\_ind, self.middle\_ind):  
 latents.append(self.styles[j](p2))  
 p1 = self.\_upsample\_add(p2, self.latlayer2(c1))  
 for j in range(self.middle\_ind, self.style\_count):  
 latents.append(self.styles[j](p1))  
 out = torch.stack(latents, dim=1)  
 return out

实验部分

## A. Dataset

我们根据不同性别和年龄的实验对象的时空头部姿势变化获取人脸图像，并使用基于 YOLO5-face 的高斯加密方法获得与原始图像配对的高斯加密数据集。为了排除图像中的其他背景特征对后续 GAN 网络高维语义抽象的影响，在人脸采集过程中，所有人脸都在同一背景下采集。下图显示了提取的部分配对数据集。这一数据加密过程也验证了视频加密方法的可行性。

## B. GANI Training and Encryption Process

我们使用 pixel2style2pixel (pSp)模型完成对上述数据集的训练。which uses the standard feature pyramid on top of ResNet to extract the feature mapping.For each of the 18 target styles, a small mapping net is trained to extract the learned styles from the corresponding feature maps. For each of the 18 target styles, a small mapping net is trained to extract the learned styles from the corresponding feature mappings. 0-2 styles are generated from small feature mappings, 3-6 styles are generated from medium feature mappings, and 7-18 styles are generated from large feature mappings. Borrowing from the human-like cognitive mechanism, corresponding to the high-dimensional abstract memory and compressed perception process of the human brain, the higher-dimensional feature information is generated from the smaller feature mappings. 其网络架构如下图所示。

我们首先从训练集中提取原始人脸 P1,使用psp encoder对P1进行训练,提取从低维到高维的多维人脸特征，并将这些信息整合加密形成密钥，与处理后的人脸身份 ID 绑定，加入人脸密钥池,存储existing\_faces.pkl中。

C. GANI 解密过程

PULSE 模型基于英伟达的 StyleGAN 算法，使用无监督图像超分辨率方法将低分辨率图像转化为高质量、高分辨率图像，从而再现肤色、眼睛、嘴唇等图像细节特征。然而，生成的高分辨率人脸图像与照片主体的真实面貌并不相似。为此，本文通过在每个生成网络层中嵌入具有正交基的子空间模型来进行半监督学习，从而获得训练模型的层次语义 B1，其网络架构如下图所示。

解密时，输入一张新捕获的人脸。使用相同的编码模型对新捕获的人脸P2进行编码提取特征，将提取到的特取某几维，跟人脸特征密钥池中的特征去匹配。根据人脸特征的相似性判断它是否在数据集中出现过。如果没有出现过，则拒绝解密。如果出现过，则确认其身份 ID，并使用之前存储的特征密钥输入到styleGAN的生成器中， 用指定 ID 的马赛克对人脸进行解密，然后输出解密后的人脸。StyleGAN 使用无监督图像超分辨率方法将低分辨率图像转化为高质量、高分辨率图像，从而再现肤色、眼睛、嘴唇等图像细节特征。

## D. Results Discussion

加密和解密效果如下所示。其中Original表示原始人脸P1,Blurred表示加密后人脸，Aged是表示对Original经过人脸老化处理的照片，本次实验中将大部分照片设置为老化到50岁，作为新捕获的人脸P2, Recoverd表示在在先前存储的人脸特征密钥的引导下还原解密以后的人脸。

1. 加密解密效果
2. 特征匹配方式

在实验中中有47张经过人脸老化处理的照片，用同样的模型提取特征后，取某几维跟人脸特征密钥池中的特征去匹配，计算匹配的准确度。实验中选择了不同的维度选取和不同的匹配标准来充分探究不同匹配方式的影响。采用的维度选取组合有从高维开始匹配的

1维，1-2维，1-3维，...，1-18维，18维，和从低维开始的18维，17-18维，16-18维，...，2-18维.采用的匹配标准有有余弦相似度(取相似度最高的)和欧拉距离(取距离最小的).不同匹配方式的匹配准确度见下表。

现在的维度选取组合有1维，1-2维，1-3维，...，1-18维，18维，17-18维，16-18维，...，2-18维（当然组合的方式还有很多种，都可以尝试；1维是最高维，18维是最低维，理论上来说，选取一些高维作为匹配依据效果比较好）

从上图可以看出，选择余弦相似度(取相似度最高的)作为匹配标准，选取一些高维特征作为匹配依据的效果更好.

|  |  |  |
| --- | --- | --- |
|  | **RLG** | |
| **Dataset** | LnAcc | IRec |
| MNIST | - | - |
| CIFAR-100 | - | - |
| ImageNet | - | - |
| ImageNet | - | - |
| CIFAR-100 | - | - |

其中，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 是大脑在头脑中根据 Ai 形成的第 i 张人脸的高维抽象语义特征集，A\_ip 是 A\_i 的加密数据集、 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低维全维人脸信息集，〖B'〗\_i是大脑在脑海中基于〖A'〗\_i形成的高维抽象语义特征集。求解上述表达式的算法如下。