**第3章 视频帧特征提取与空间变换**

3.1. 图像预处理

由于原始采集的图像存在亮度较低，细节不明显等问题，所以必须对原始图像进行预处理。本论文的主要图像预处理方法为灰度等级变换。

3.1.1. 灰度等级变换原理

图像的灰度等级变换与像素在图像中的位置无关。定义变换，将原来图像灰度等级内的亮度*p*变换为一个新的范围内的*q*，其定义如下：

(2.1)

图3.1展示了常见的灰度等级变换；分段函数a增强了图像在亮度和之间图像的对比度。函数b被称为亮度阈值化，其结果是二值图像。函数c是负片变换。数字图像的灰度级别是离散的有限集合，因此通过灰度查找表将十分容易实现灰度的等级变换。该原理同样容易在彩色图像上实现，对与彩色图像可以分解为在不同色彩空间下的灰度图，在不同色彩空间下进行相同的灰度等级变换，就可以实现对彩色图像的灰度等级变换。

255

255

a

b

c

*p*

*q*

图3.1. 灰度等级变换常用函数

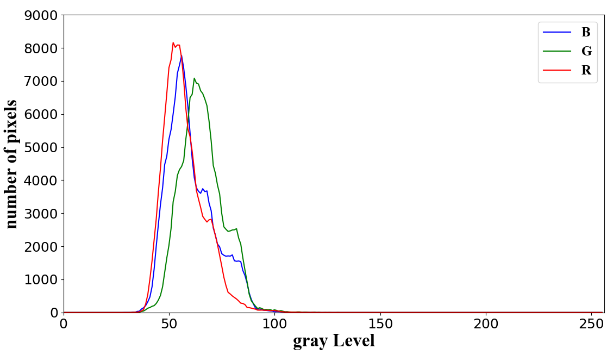
根据第二章对UVDASSB的图像特点分析，该数据库下的图像（视频帧）灰度等级较少并且亮度较低，所以这里采用的灰度变换函数为函数a，更加具体的说，这里的图像预处理方法为线性变换，本文所采取的灰度变换方法可以定义为如下的公式

式中，为线性变换系数；*p*为原始图像下的灰度等级；*q*为经过灰度等级变换后的灰度等级。

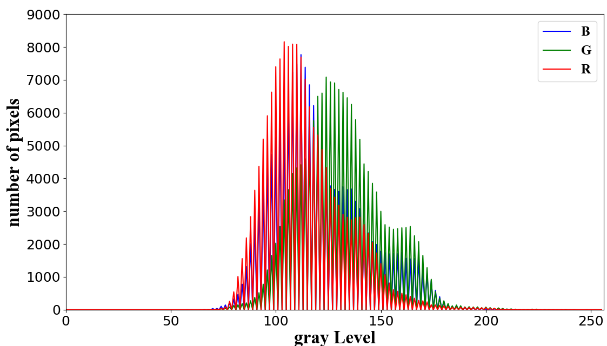
3.1.2. 图像预处理结果

通过对UVDASSB中视频样本的灰度等级的统计我们发现96%的灰度等级均低于120，于是这里的线性变换系数被设置为2。在线性变换过程中，视频的RGB三个颜色通道都进行了相同的线性变换。因为90%的像素灰度值都低于120，经过线性变换后最大的像素灰度值为240低于255。但是存在部分的灰度值经过线性变换后大于255，我们直接将其设为255。

根据如上的灰度线性变换，最后得到的实验结果如图3.2所示，a图为未经过预处理的灰度图像，b图为对应的灰度直方图，其中不同颜色的曲线代表不同的颜色空间灰度直方图；c图为线性变换后的图像，d图为对应的直方图。通过实验结果可以得出线性变换的主要两个功能：第一，通过线性变换提高了图像的灰度等级范围，原来的图像灰度范围为50-100，只有50个灰度等级，而经过线性变换后灰度等级范围变为70-200，有130个灰度等级。第二，线性变换后的图像亮度得到了提高，原来的平均灰度等级为70而经过线性变换后得到的平均灰度等级为140。



(a) (b)



(c) (d)

图3.2 图像预处理结果（a）预处理前图像（b）预处理前灰度直方图（c）图像预处理后（d）图像预处理后灰度直方图

3.2. 变分贝叶斯推断与图像变分自编码器

我们假设图像尺寸为并且为彩色图像，则从概率学的角度该图像可以被视为从一个维的图像空间下的一个观测结果，概率记作。显然求取这样一个超高维度的概率是十分困难的，并且在这个超高维空间下相当一部分的变量在我们的观测空间下是没有意义的。比如当这个空间下所有维度的分量都为0的时候，则观测结果为一副全黑的图像，这样一副全黑的图像在一般的数据集下都是没有任何实际意义的，所以我们可以认为这样的观测结果出现的概率是极低的。除了0以外，大部分的随机出现的各种维度的变量得到的结果都是杂乱无章的数据图，如图3.3我们随机生成了一个维的矩阵，得到结果是毫无意义的杂乱无章的矩阵，在一般的数据集下通常得不到这样的观测图。

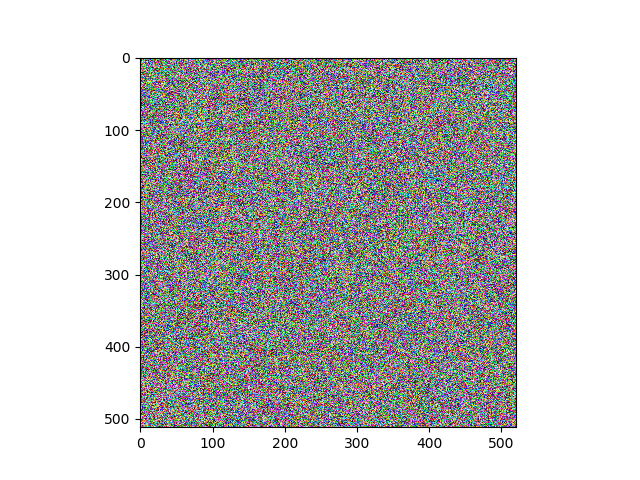
**

图3.3 在512\*512\*3高维空间下随机采样图像

在一般数据集下的图像样本实际上是按照一定规则在超高维空间下进行观测，其观测结果与图像拍摄场景和拍摄角度有关。更具体的说，观测的过程是在有约束和条件的在高维空间下进行观测。于是我们将图像的观测过程描述为两个步骤，首先在某个隐含空间下进行条件观察，而在图像空间下的观测是基于隐含空间下的观测结果进行的，其概率图模型如图3.4所示。隐含空间被定义为z,而图像的观测结果定义为x。图像的生成过程可以描述如下，首先在隐藏空间下生成隐含空间变量其概率可以表示为，在隐变量的基础下得到的图像空间的概率可以表示为,所以最后的生成式模型表达为图3.4中的实线则表示生成式模型。而图中的虚线为对隐含空间的推理，当已知图像空间下观测结果为x时，则隐含空间变量为z的概率。显然这个概率的表达形式以及概率特征都是及其难以获取的，这里我们用一个简单的概率去近似的求取真实的后验概率，这个简单的后验概率被定义为，其概率密度函数形式为卷积神经网络，整个近似优化的过程可以通过变分推断求取。

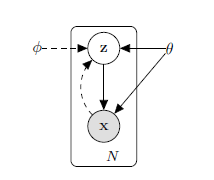


图3.4. 概率观测模型

小节3.2将会介绍变分推断以及变分自编码器的基本原理，小节3.3通过变分自动编码器将UVDASSB数据集下的视频数据进行空间变换，原始视频图像将被映射到隐含空间。通过空间变换可以达到两个目的：数据压缩和将图像 映射为概率分布。

3.2.1. 变分贝叶斯推断

通过变分推断，一个复杂分布的概率密度函数可以通过某个已知其概率密度函数结构近似，在不同的变量情况下调整已知概率函数的参数使得其与未知的概率密度函数最接近。在变分推断中，这个近似问题被转换成了一个优化问题。

首先定义未知的概率分布为,被用于近似的概率为。如何实现通过近似p描述如下。

根据恒等变换我们有如下公式，

等式右边的第一部分为期望，第二部分为散度的定义，所以有，

式中为变分推理下界，记作ELBO, 为形式已知的后验概率与复杂的未知后验概率分布的散度距离，记作KL。

回归我们的目的，利用一个已知形式的简单分布去近似复杂的难以求取的概率分布，实际上就是最小化KL，当KL等于0时则两个概率分布完全相同。直接优化散度距离是困难的，因为这里的未知，但是通过公式（3.1）可以实现问题的转换，公式左边与q(x)无关所以当调整q(x)的参数时公式左边为一个常数，又因为KL>0,所以最小化散度距离的问题转换成了最大化ELOB的问题，其示意图如3.5所示。显然，通过调整q(z|x)的参数，当ELOB越大则KL越小，当ELOB取得极大值的时候则KL取得极小值。

**调整**

图3.5 变分推理优化散度距离原理示意图

下面介绍优化ELBO以使得KL最小。

3.2）

因为散度距离恒大于等于0，则为使得ELOB最大则需要最小化公式3.2第一项中的散度距离和最大化第二项的期望。在下一小结我们将介绍基于此的图像编码压缩和图像重建。

3.2.2. 变分自动编码器

变分贝叶斯推断可以将对隐藏空间推理的复杂条件概率分布 通过一个卷积神经网络近似，这个卷积神经网络被定义为编码器（Encoder）。当编码器实现了对的近似则原来的图像可以表达为隐含空间的后验概率分布。而当隐含空间变量已知时的后验概率被认为是一个协方差矩阵为对角矩阵的多元高斯分布（MGD），这个MDG用一个逆卷积网络表示，并且被定义为解码器（Decoder）。

1. 编码器

在变分自动编码器中，编码器的主要目的是将输入图像从图像空间映射到隐含空间下的MGD。在贝叶斯变分推断中，编码器即，通过对公式3.2的优化当ELBO为最大时，则与的散度距离最小，即成功实现了对的近似。

由于的形式可以任意设置，在变分自动编码器中的形式为协方差矩阵为对角矩阵的MGD，并且由一个卷积神经网络表征，其形式如下，

其中，

tanh为非线性响应函数，flatten为将矩阵拉直操作，convs为多层的卷积神经网络，编码器示意图如图3.6所示。

X

convs

h

图3.6 变分自动编码器解码器部分

1. 重参数技巧

通过解码器得到的图片在隐含空间下的MGD，由于需要对公式3.2进行优化，而直接通过图像在隐空间下的概率分布是难以对进行优化。部分的优化采用了蒙特卡洛采样优化算法，即

其中，为以为随机概率的抽样结果，L为抽样次数。

在变分自动编码器的优化中一般选择的是随机梯度下降法，为使得网络能够反向传播，在变分自动编码器运用了重参数技巧，其网络结构可以表述为图3.7。

**+**

**\***

Sample from

Latent space code

图 3.7 重参数网络结构

即通过重参数后得到的隐藏空间编码向量为，

式中，为从的随机采样值，与为编码器输出值。

1. 解码器

形象的解释解码器的主要功能为将隐含空间下的编码复原为一张图像。而从概率的角度分析实际上是在已知图像在隐含空间下的编码，解码器能以最大的概率使得该编码可以映射为图像空间下的原图像。现假设我们已经拍摄了一张以计算机为主要内容的图像，被映射大隐藏空间后其编码信息为“图像主体为计算机”，基于这样的隐藏信息解码器能以最大的概率重建一张以计算机为主题的图像，但是值得注意的是这张图像与原图像是存在差异的，差异的大小取决于隐含空间包含信息的多少，如果隐含空间下的信息为“图像主体为计算机正面”，那么解码器的重建的图像与原图像差异更小。

编码器优化数学表达如下：

其中p(x|z)为已知隐含空间编码的后验概率，且用一个MGD进行表示，而MGD的表达形式为一个逆卷积神经网络表述如下：

其中，

根据公式3.2，优化目标为使得以函数的期望最大与方差无关，所以在网络结构中可以略去对方差的重建，deconvs1为逆卷积函数，tanh为非线性函数，reshape为将向量重排为图像矩阵，dense为对隐含空间下的编码进行全连接层操作。解码器网络结构示意图如图3.8.

dconvs1

im

图3.8 解码器网络结构

1. 损失函数

根据前述编码器与解码器的设计，公式3.2优化目标可以描述为如下两部分：

对part1的优化实际上是调整编码器的参数，即优化编码器使得图像在隐藏空间的后验概率分布（）与隐藏空间的先验分布（）散度距离最下。隐含空间的先验分布在变分自动编码器中被定义为标准多元高斯分布（实际上这是一个先验假设，也可以假设为其他分布，这里假设为标准MGD），而可以被定义为任何一个简单的分布，这里同样被定义为MGD。所以对part1的优化变为了求取两个MGD的散度距离，具体的计算过程可以参见附录1，将为标准MGD，得到下式：

式中，*q*为编码器；J为隐藏空间的维度；与分别为第j个维度分量的方差与均值。

而对公式3.2优化的第二部分可以描述为如下：

通过蒙特卡洛采样的方法part2的优化目标可以转换为对图像空间下后验概率的最大优化，在变分自动编码器图像空间的分布被近似定义为MGD，所以当它的期望与输入图像相同时其概率最大，其数学表述如下：

式中，p为编码器；为重参数采样值；为编码器输入图像，*x*为解码器的输出值。

综上，变分自动编码器的损失函数可以定义为如下：

1. UVDASSB数据实验

为验证变分自动编码器在UVDASSB下的图像编码与重建功能，变分自动编码器在UVDASSB的图像数据集下进行了训练，编码器卷积神经网络参数设计如表3.1所示，其中包含了三个卷积层（Cov1, Cov2, Cov3），Cov1的卷积核尺寸为5\*5，卷积步长为1，输出通道为16，Cov2卷积核尺寸为5\*5，卷积步长为1，输出通道为32，Cov3卷积核尺寸为5\*5，卷积步长为1，输出通道为64；一个池化层（Pool1），池化类型为平均值池化，池化最小单元为2\*3；一个图像拉升操作（Flatten）将图像矩阵转换成成向量；以及一个全连接层（FC），FC的输出向量为50。编码器输入为304\*540\*3图像，输出为

表3.1 变分自动编码器编码器网络参数

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 名称 | Cov1 | Cov2 | Pool1 | Cov3 | Flatten | FC |
| 参数 | 5\*5\*16 | 5\*5\*32 | 2\*3 | 5\*5\*64 | \ | 50 |

解码器逆卷积神经网络如表3.2所示，其中包含了一个全连接网络（FC），输出向量长度为27986；图像变形操作（Reshape）将FC层的输出向量变形为大小为19\*23，通道数为64的图像；三个逆卷积层（Decov1, Decov2, Decov3），Decov1的逆卷积核尺寸为5\*5，逆卷积步长为1，输出通道为32，Decov2的逆卷积核尺寸为5\*5，逆卷积步长为1，输出通道为64，Decov3的逆卷积核尺寸为5\*5，逆卷积步长为1，输出通道数为3；一个池化层（Unpool1），其逆池化最小单元尺寸为2\*3，采用直接填充方式；

表3.2 变分自动编码器解码器网络参数

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 名称 | FC | Reshape | Decov1 | Unpool1 | Decov2 | Decov3 |
| 参数 | 27986 | \ | 5\*5\*32 | 2\*3 | 5\*5\*64 | 5\*5\*3 |

完整变分自动编码器的结构框图如图3.9所示，UVDASSB数据库中的图片通过编码器被编码为隐藏空间MGD并解码为图像。

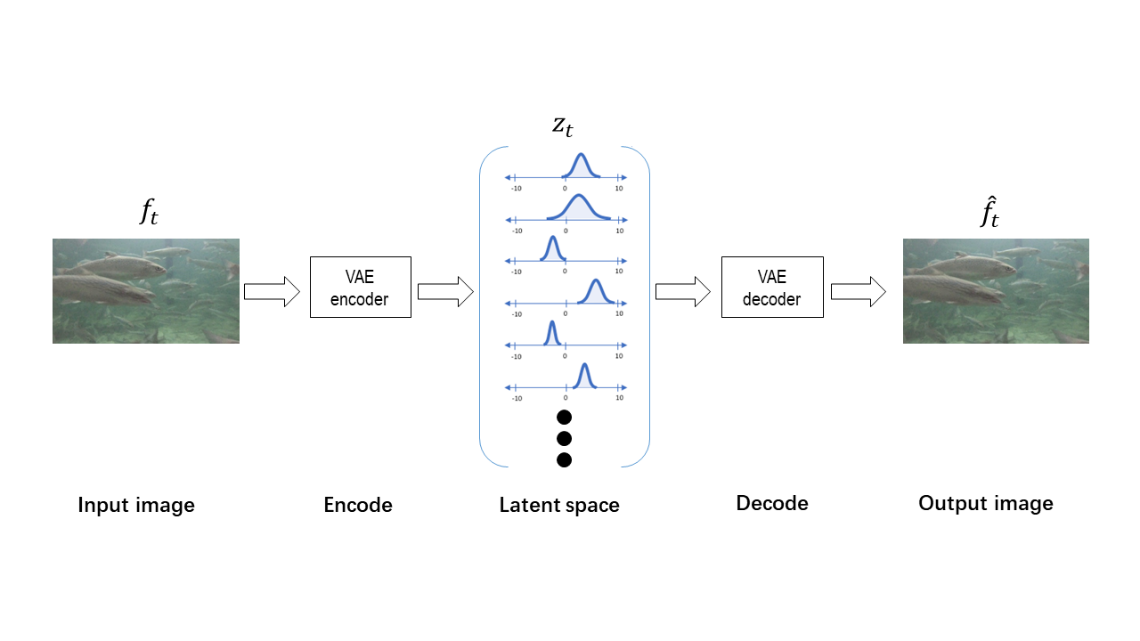


图3.9 变分自动编码器完整框图

网络训练损失函数为公式3.3，训练损失函数图如图3.10所示。分析图3.10显然可以看出VAE网络在1000次迭代后基本收敛，同时还应该注意到相比其他的数据集，在UVDASSB下训练的VAE在前10次的迭代过程中损失函数的下降速度非常的快。对于这种异常的状况可能有两种可能：一、训练集过小，数据快速过拟合；二、数据集样本情况特殊。对于第一种情况我们可以通过验证集的损失函数进行检验，可以看到在图3.10中的验证集损失函数值与训练集的测试函数值相差不大所以不会是过拟合问题。实际上造成这种情况的原因正是由于数据集情况特殊。在UVDASSB中，数据的主题内容是鱼类的形状与姿态，而所有图像的背景都是水环境，而且环境状态基本一致，所以图像整体相似度很高（图3.11）。

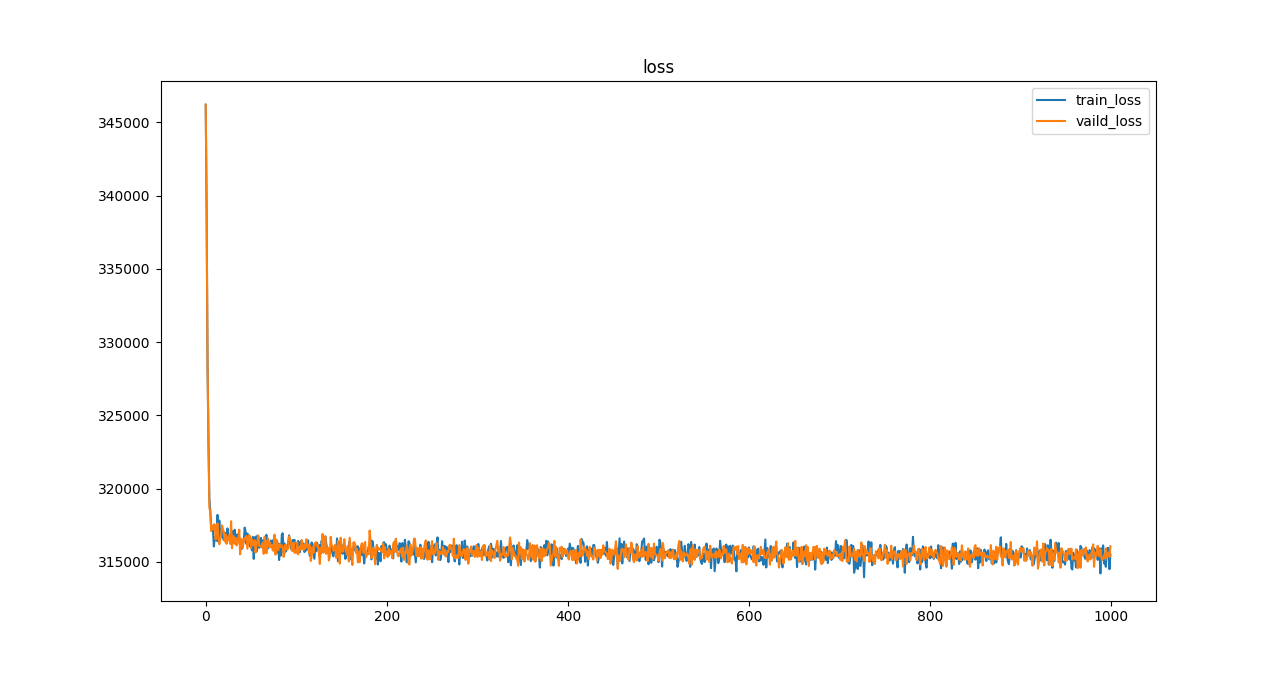


图3.10 VAE训练损失函数图

53

48

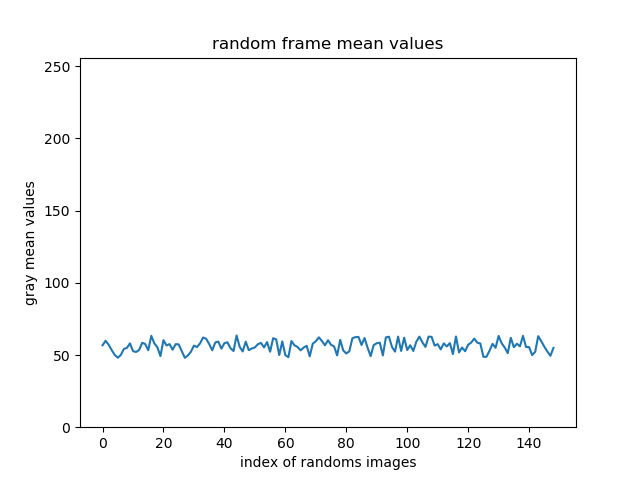


图3.11 在UVDASSB中随机选取150张图计算灰度均值，可以看到均值波动在48到53之间，说图像在整体范围下的变动不大，具有一定的相似度。



（a）



(b)

图3.12 UVDASSB数据库原图（a）与训练后数据重建图（b）

训练完成的VAE网络实现了对UVDASSB图像数据的编码与解码其实验结果如图3.12所示。图3.12a与3.12b各自包含了50张图像，3.12a为UVDASSB数据库中的原图，3.12b为将UVBASSB送入VAE编码然后通过解码器复原得到的重建图像。通过对比原 图像与最后的重建图像可以看出，从隐藏空间重建得到的图像与输入图像有差异但是重建图像基本恢复了鱼类的形状与姿态，这对于本文的鱼类摄食行为模型研究是非常重要的，因为鱼类摄食行为的主要研究对象就是鱼类的形状与姿态。于是我们可以得出结论，基于VAE编码后隐藏空间的MGD包含了鱼类摄食行为的形状与姿态，可以满足对鱼类摄食行为模型研究的需要。

3.3. UVDASSB数据空间变换

在目前大多数的研究都集中于VAE的解码网络部分（Decoder），基于解码网络可以生成新数据。但是VAE的解码网络却经常被研究人员忽视，实际上编码 网络具有两个重大的功能：（a）.数据压缩，编码网络实现数据压缩是一个很有效的在给定数据集下的数据压缩，尤其在处理视频数据时一个一分钟的视频可能含有数千张的图像，那么直接对这数千张的图像进行计算则将占用大量内存，计算量巨大。通过变分自动编码器的编码网络则原始的图像被转换为了两个向量，向量的维度等于隐含空间的维度。（b）.提高数据泛化功能，在视频模型的建立过程中，我们往往无法建立严格精准的模型，所以在一般的视频建模中我需要人为的引进噪声项而避免模型无法训练的后果，但是基于变分自动编码的编码网络，每一张图像对应的两个向量分别代表均值向量和方差向量，其中均值向量表明该图像在隐含空间下的方向以及大小，而方差向量则表明在该向量在其均值附近的波动范围。由于方差向量的存在，使得原本固定的图像成为了在隐含空间中按照方差和均值变量随机采样的一个向量。这意味着，运用这些数据训练模型之前数据已经进行了一次泛化，因此无须再额外加上增强模型泛化和鲁棒性的噪声项。

本小节所提出的数据空间变换就是利用变分自动编码器的编码器部分将视频数据进行空间映射，将像素空间下的视频帧数据映射到隐藏空间下的MGD。在UVDASSB中每个样本视频片段的长度为5秒，150帧图像（fps为30），图像原本的尺寸为304\*540；隐藏空间下的MGD维度为超参数，在本论文中被设置为50。空间映射的实验结果如图3.13所示，视频数据样本被映射为两个矩阵，分别为多元高斯均值矩阵与多元高斯方差矩阵，这两个矩阵的行数与隐藏空间维度相等，在本论文中为50，列数与视频样本的帧序列长度相等，在本论文为150。

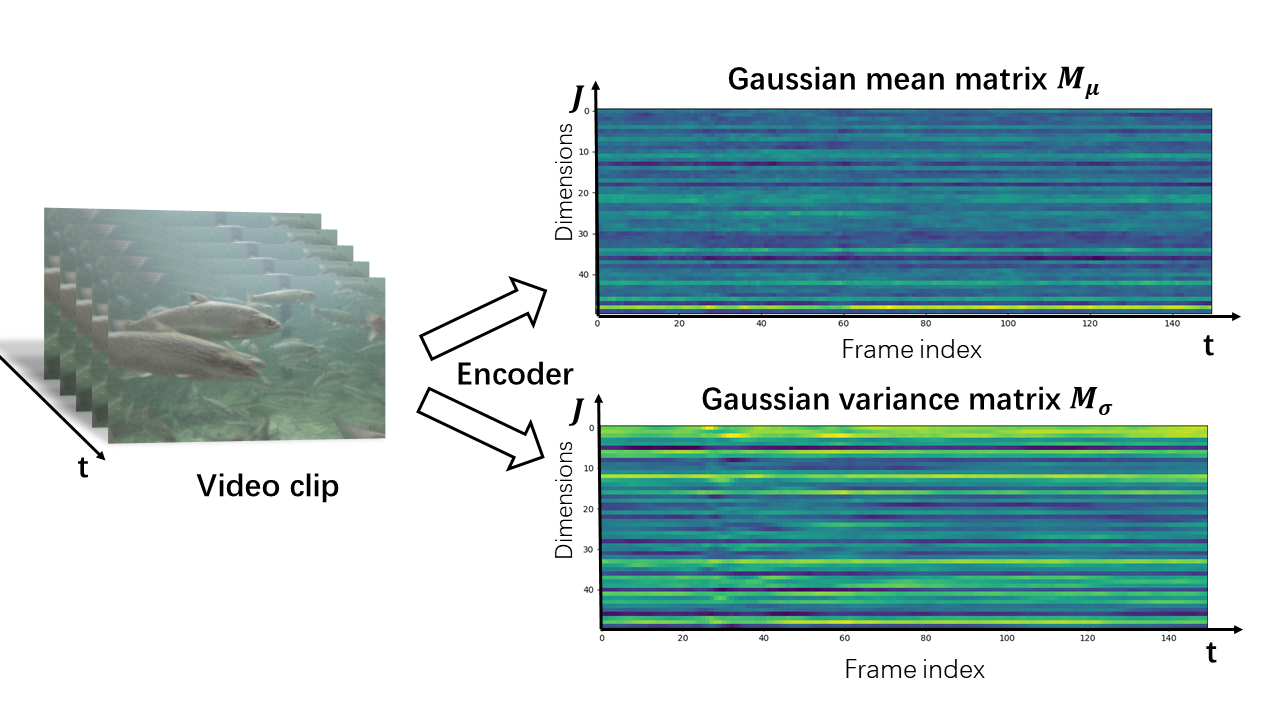


图3.13 视频数据空间变换

为评估通过这样的空间变换对视频数据压缩的贡献，我们采用压缩比进行评价，压缩比定义如下：

所以基于VAE编码器的图像压缩比为：

式中，*f*为视频样本帧数，*J*为隐藏空间多元高斯维度，*w*为图像宽度，*h*为图像高度。

将*w=*540，*h*=304， *J*=50代入其中可以得到压缩比为1：32832‬。将UVDASSB中所有的视频样本进行压缩后并通过序列化后数据的实际存储大小为，摄食视频样本：37MB ；未摄食样本：179MB 详细的视频映射前后对比参见表3.3。

表3.3 视频数据空间映射前后对比

|  |  |  |
| --- | --- | --- |
|  | 像素空间下 | 隐藏空间下 |
| 视频数据形式 | 连续视频帧 | 两个矩阵 |
| 视频数据大小 | 304\*540\*3\*150 | 50\*150\*2 |
| 视频数据特点 | 细节丰富、清晰 | 数据具有随机分布特点 |
| 优点 | 准确描述场景 | 泛化性好、数据量小 |
| 缺点 | 处理计算大、冗余高 | 有损压缩、与场景有误差 |

必须注意到，基于变分自动编码器的数据空间映射是一种有损的数据压缩方式，只能在特定的场景中运用。比如当我们对鱼类的摄食行为模型进行研究时，研究的主要内容是鱼类的形状与姿态，与水体背景相关性不大，而隐藏空间的高斯均值矩阵和高斯方差矩阵通过解码器是可以重建鱼类的形状和姿态的，所以使用隐藏空间的矩阵研究是充要的。

3.4. 本章小结

本章的主要内容是基于UVDASSB的数据进行分析与处理。首先由于水下光环境的情况特殊，我们选择了线性变换的图像预处理，提高了图像的亮度与灰度等级，为进一步的处理奠定了基础。完成了对数据的预处理后，从变分贝叶斯推断基本原理出发介绍了变分自动编码器。利用变分自动编码器的编码功能实现了视频数据的空间变换。

视频数据的数据空间变换的目的主要有两个：一、数据压缩，便捷的发布UVDASSB，通过空间变换后的视频数据片段，在隐含空间中仍然包含鱼类的形状与姿态，是瞒住鱼类摄食行为模型研究的，而空间变换可以极大的压缩数据集，便于网络传输与发布；二、视频模型分类，在下一章本论文将介绍基于隐藏空间的均值矩阵与方差矩阵实现鱼类摄食行为分类的模型。