**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

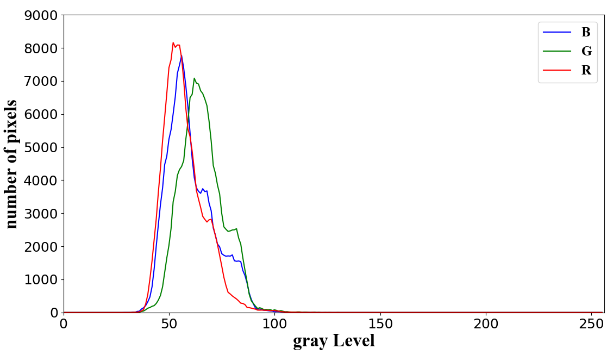
根据第二章对UVDASSB的图像特点分析，该数据库下的图像（视频帧）灰度等级较少并且亮度较低，所以这里采用的灰度变换函数为函数a，更加具体的说，这里的图像预处理方法为线性变换，本文所采取的灰度变换方法可以定义为如下的公式

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。

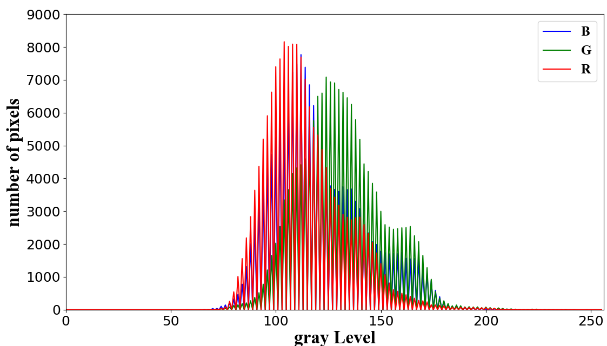
3.1.2. 图像预处理结果

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。

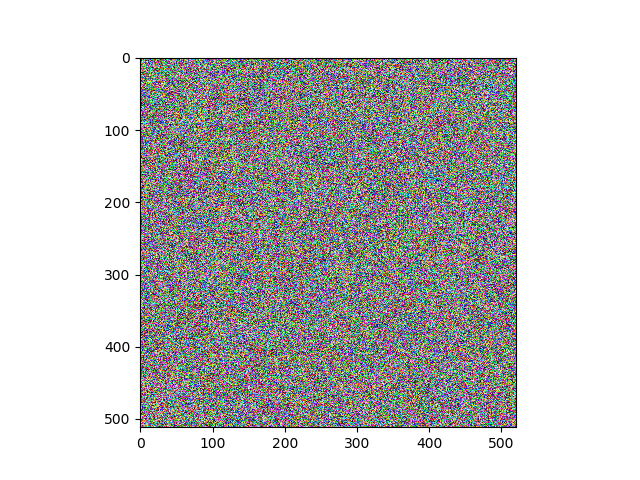
**

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。

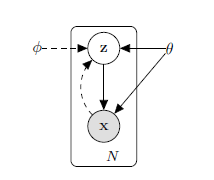


图3.4. 概率观测模型

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：数据压缩和将固定图片映射为概率分布。

3.2.1. 变分自动编码器

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。

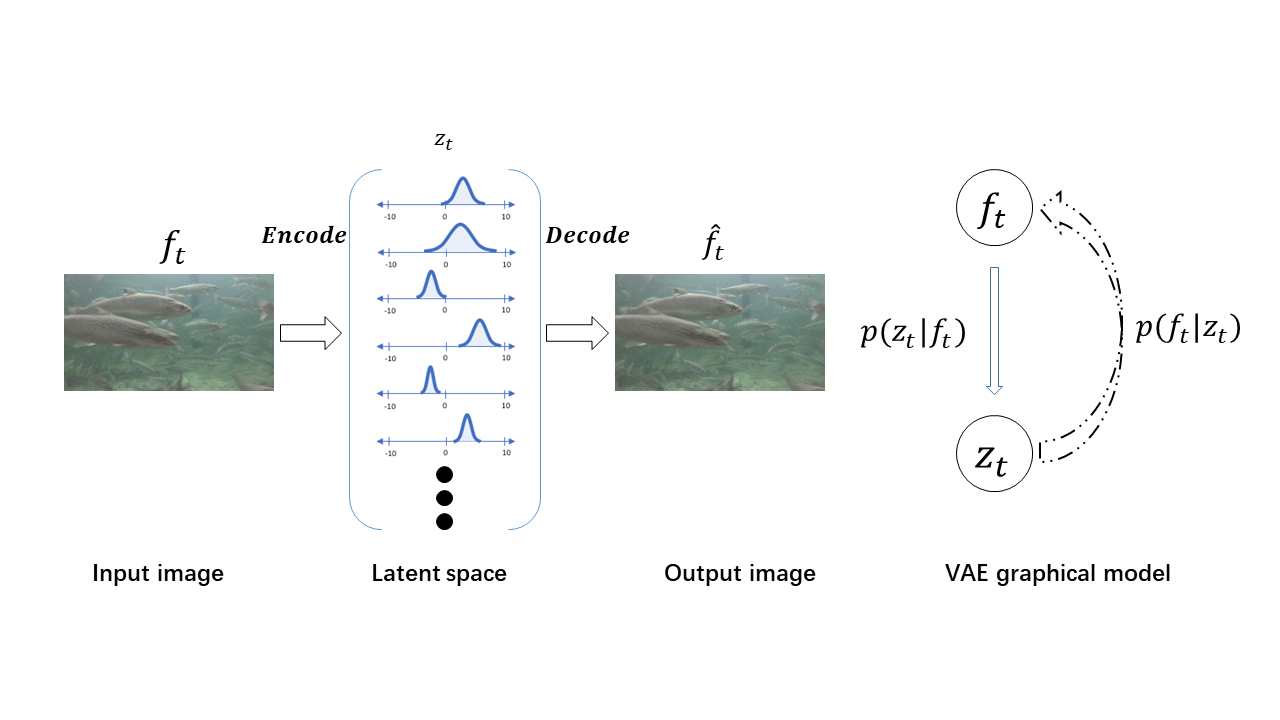
3.2）

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，如何通过卷积神经网络实现对这两项的优化。

3.2.2. 卷积神经网络近似

为实现对公式3.2的优化，则必须先找到公式中变量的合理表达，首先是,在变分自动编码器中，隐含空间p(z)被定义为标准正态分布，空间维度为J，则;然后是当图像已知时隐含空间的后验概率的近似概率，在变分推断中，我们希望通过某个简单的近似真实的,这里我们定义的形式为卷积神经网络，且服从正态分布，更具体的说的均值与方差通过一个卷积神经网络得到，神经网络的输入为观测空间得到的x，输出为正态分布的均值与方差，这里的第一个卷积神经网络被定义为编码器（encoder）；最后一个变量为，在已知隐含空间时，观测图像的后验概率分布，根据文献【48】，的分布可以被定义为高斯分布或者伯努利分布。当观测图像为二值图像时，例如MINIST【49】数据集下的手写数字图像可以通过一个伯努利分布表示，而对非二值图像则通过高斯分布来表示。因为本文的数据是灰度图像，所以这里的为高斯分布，与编码器相同的均值与方差通过一个逆卷积神经网络生成，这里的逆卷积神经网络被定义为解码器（decoder）。

3.3. 视频帧空间变换

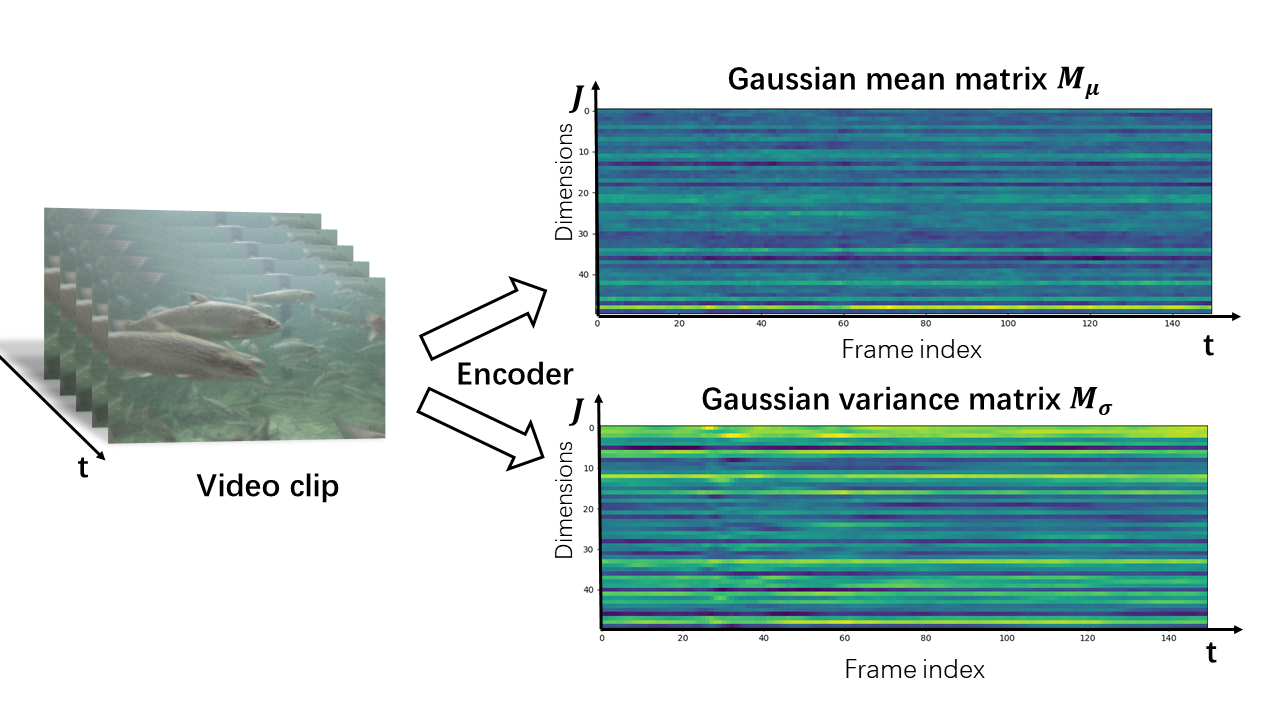


3.3.1. 变分自动编码器训练

3.3.2. 视频帧空间变换

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，实际上解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的噪声项。

（1）实验结果



（2）实验结果分析

3.4. 本章小结