

# Time Series To Images: Monitoring the condition of industrial assets with deep learning image processing algorithms

论文地址: <https://arxiv.org/abs/2005.07031>

许多应用领域中, 基于时间序列检测异常的功能被认为具有很高研究价值。时间序列对象自带的顺序性造成了额外的功能复杂性, 导致需要专门的方法来解决任务。当未对时间序列进行任何变换时, 通常无法使用最新的异常检测方法来捕获时间序列时间领域之外的基本特征。受到深度学习方法在计算机视觉中成功的启发, 一些研究提出将时间序列转换为图像形式的表示。但是, 大多数以前对图像编码执行时间序列的研究都集中在监督分类上。无监督异常检测任务的应用受到限制。

## Literature Review

随着许多工业市场的全球化, 对于运营商而言, 优化其资产的使用和运营成本变得越来越重要。这种优化要求能够在法律规定的安全级别上保持高可用性。由于工业资产的维护占了运营成本的很大一部分, 因此能够监视系统并根据其状况进行维护将极大地帮助操作员。

随着最近传感器成本的下降以及将大型监控环境与工业物联网连接的可能性, 如今捕获大型数据流已成为可能, 即使对于高频数据也可以实时访问。深度学习的最新发展提供了从大型数据集自动学习签名的可能性, 这将最大化其训练目标。

如今, 深度学习可在受监督的setup中提供最先进的结果。也就是说, 在对算法进行训练的前提下, 要对所有条件进行足够的代表性采样以区分。然而, 异常检测是一项更具挑战性的任务, 因为异常仅针对主要类别定义, 而不是针对自身定义。在该行业中, 不可能定义并因此收集所有可能类型的异常的代表性样本。因此, 需要设计一种用几乎完全在正常条件下收集的数据进行训练的异常检测器。

实际生活中, 复杂的工业系统几乎只能在正常的条件下运行, 否则要对其进行维护以避免严重后果。正常操作数据通常是高度可用的, 因此, 异常检测的任务可以定义为学习正常操作条件的表示形式以检测是否存在偏差。以前的工作表明, 对于这种方法, 监视在主类上训练的自编码器的残差可以解决此问题。然而, 深度学习的最新发展受到图像处理的推动, 在自动编码器的使用方面取得了一些重大成功。此外, 因为对图像进行时间序列编码可以帮助强调, 捕获或压缩本来会随时间分布的局部模式。(In addition, encoding time series to images can help to emphasize, capture or condense local patterns that would otherwise be spread over time.) 最近的工作建议将时间序列转换为图像。

## Contribution

本文具有以下贡献:

- 1、首先, 论文评估了六个时间序列在DL算法的图像编码中的应用: **GAF**格拉姆角场, **MTF**马尔可夫转移场, **RP**递归图, **GSE**灰度编码, **Spectrogram**频谱图和**Scalogram**比例图。
- 2、其次, 论文提出了对原始编码定义的修改, 以使其表现对大型数据集的可变性更加健壮。

3、第三，论文提供了直接使用原始时间序列和使用不同编码（有无建议的改进）之间的全面比较。这些比较是基于Airbus SAS 2018收集并发布的数据集上进行的，其中包含来自实际直升机飞行测试的高度复杂的振动测量结果。不同的编码为异常检测提供了有竞争力的结果。

## Framework for Anomaly Detection

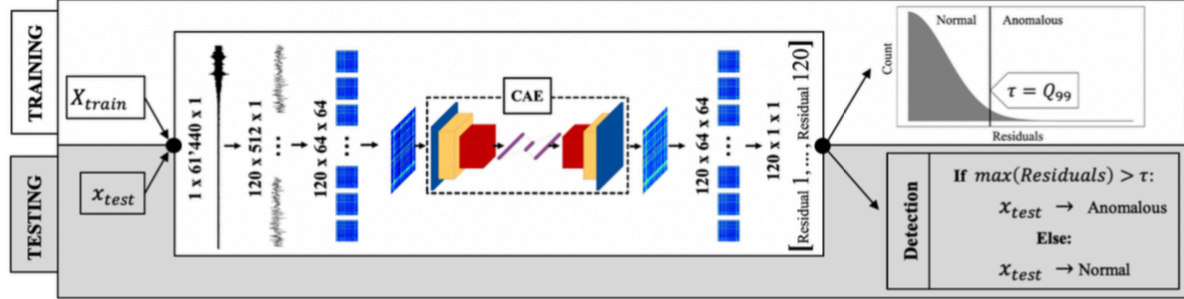


Figure 1: **Framework.** The network is trained on  $N$  time series in the training data set  $X_{train}$  and the threshold is set as the 99-th percentile over the sample of  $120 \times N$  residuals ( $108 \times N$  for the GS encoding). For testing an arbitrary test sample  $x_{test}$ , each of the 120 slices (108 for the GS encoding) is reconstructed by the CAE and the largest residual is compared with the threshold  $\tau$  to detect anomalies.. When using the raw time series, the same framework is followed without the encoding step and with 1-D convolutions in the auto-encoder.

Figure 1中描述的框架首先包括时间序列拆分，其次是每个子序列的编码，最后是在通过卷积自动编码器（CAE）馈送时监视其聚合残差。

在训练数据集  $X_{train}$  中对  $N$  个时间序列进行网络训练，并将阈值设置为  $120 \times N$  残差样本（对于GS编码为  $108 \times N$ ）的**第99个百分位数**。为了测试任意测试样本  $x_{test}$ ，通过CAE重构了120个切片中的每个切片（对于GS编码为108个），并将最大残差与阈值  $\tau$  进行比较以检测异常。

我们将结果与类似的框架进行比较，以原始子系列作为输入，并使用一维CNN。

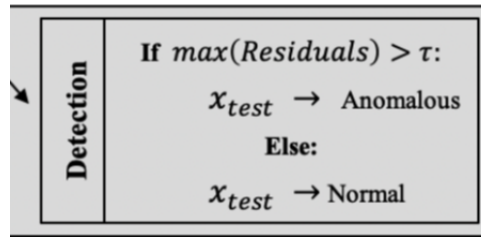
论文提出的框架如图1所示。包括基于自动编码器神经网络的异常输入重构检测，该神经网络在**时间序列上将不同类型的图像编码作为输入**。基本思想是训练自动编码器以重构其自身的输入，同时将其压缩到较低维的潜在空间中。网络将按照与经过训练的分布相同的分布，**学会尽可能地重建其输入数据**（时间序列或图像）。如果在正常数据上训练自动编码器，则如果呈现出偏离训练过程中使用的正常状况的异常状况，则说明其重构能力将很差。监测与阈值有关的重建误差可以检测到这种异常模式。

因为数据集中的**原始时间序列相当长**（数千个值），因此论文提出了**两种挑战**。首先，由于监视的残差是整个重建过程的总和，因此局部异常在全局范围内可能并不明显。其次，对于大多数编码，计算时间和长短会根据输入大小增加而指数增长。为了克服这两个挑战，我们建议首先将时间序列细分为大小为  $l_1$  的较小子序列。然后将这些子序列编码为大小为  $l_2 \times l_2$  的图像，这些图像由神经网络重建，并计算和汇总其残差。我们将第  $k$  个切片的残差  $Res_k$  定义为大小为  $l$  的输入  $x_{or}$  与其重构  $x_{rc}$  之间的差的  $l_1$ -范数：

$$Res_k = \sum_{i=1}^l (|x_{or}^i - x_{rc}^i|), \quad (1)$$

baseline采用的是没有任何编码的原始子序列1-D数据输入CNN，其中长度  $l = l_1$ ；其余情况采用的输入是  $l = l_2 \times l_2$ 。

论文仅使用正常状况下的时间序列进行训练。基于所有子序列的残差分布，我们提取第99个百分位数以定义检测阈值 $\tau$ 。在测试集中，对于每个时间序列，我们监视其子序列上的最大残差，并将其与阈值 $\tau$ 进行比较。如果超过阈值，则将时间序列检测为异常。监测残差的最大值使得能够检测将影响一个或几个子序列的局部异常。



挑战一：时间序列长度过长——》切割时间序列

挑战二：无法顾及全局范围的时间序列影响——》检测子序列的最大残差

For consistency with the framework designed in Figure 1, we split the time series of length 61 440 into 120 slices of size  $l_1$  equal to 512. No additional operations such as smoothing or normalization is performed.

## Methodology

### Time Series To Image Encoding

#### Gramian Angular Field (GAF)

格莱姆角场由两步过程组成：首先，执行从空间（时间×值）到极坐标的转换。然后，计算该新空间中的Gramian矩阵。转换为极坐标包括执行以下操作：

$$\Theta_i = \arccos(x_i)$$

$$r_i = \frac{i}{L}$$

其中 $L$ 缩放时间维度，但不影响最终矩阵。这种转换是双射的，并且保留了绝对的时间关系（Wang and Oates 2015）。由于方程式2仅针对 $X \in [-1, 1]$ 定义，因此首先需要缩放数据。Wang和Oates 2015建议将输入 $X$ 扩展为：

Given a time series  $X = \{x_1, x_2, \dots, x_n\}$  of  $n$  real-valued observations, we rescale  $X$  so that all values fall in the interval  $[-1, 1]$ :

$$\tilde{x}_i = \frac{(x_i - \max(X)) + (x_i - \min(X))}{\max(X) - \min(X)} \quad (1)$$

最后Gramian matrix可以用点积去计算：

$$G = \begin{pmatrix} \cos(\phi_1 + \phi_1) & \cos(\phi_1 + \phi_2) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cos(\phi_2 + \phi_2) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cos(\phi_n + \phi_2) & \cdots & \cos(\phi_n + \phi_n) \end{pmatrix}$$

类似于Gram矩阵  $G_{ij} = \langle x_i x_j \rangle$ ，其中  $\langle \cdot, \cdot \rangle$  为内积，其矩阵名为Gramian Angular Field。但是，可以注意到， $\cos(x_i + x_j)$  不能满足内积的所有条件。另外，应注意，矩阵不依赖于  $r$ 。因此，它不依赖于时间，也不依赖于等式2中  $L$  (时间序列长度) 的选择。因此，**GAF变换不可逆，这意味着GAF不允许对时间序列进行唯一的重构。在应用编码后，这导致信息丢失。**然而，由于在逆空间中通常只有少量的重叠，所以输入可以大致近似看作一一对应的映射。

**论文提出的改进：**在我们的研究中，我们建议根据完整的训练数据集来缩放训练样本，以保持样本之间的关系和差异。因此，我们建议根据训练集分布来定义  $UB$  (upper bound 上限) 和  $LB$  (lower bound 下限)。如果在测试集中样本中包含的值超出了这些界限，我们建议将这些值裁剪到界限值。

## Markov Transition Field (MTF)

马尔可夫转移场包括建立概率矩阵，以观察时间序列中任何一对点之间的值变化。首先，时间序列是离散的。然后，在整个训练集中，每两个连续的数据点对从一个仓bin到另一个仓bin的转变  $w$  进行计数并归一化。最后，MTF矩阵针对时间序列中的每对点包含点所属的两个仓之间的转移概率。

$$M = \begin{bmatrix} w_{ij}|x_1 \in q_i, x_1 \in q_j & \cdots & w_{ij}|x_1 \in q_i, x_n \in q_j \\ w_{ij}|x_2 \in q_i, x_1 \in q_j & \cdots & w_{ij}|x_2 \in q_i, x_n \in q_j \\ \vdots & \ddots & \vdots \\ w_{ij}|x_n \in q_i, x_1 \in q_j & \cdots & w_{ij}|x_n \in q_i, x_n \in q_j \end{bmatrix}$$

时间序列的这种转换是不可逆的，这意味着存在大量信息丢失。特别是，时间序列的离散化在转换中保留或丢失的信息量中起着至关重要的作用。设置大量仓bin可能会导致图像稀疏，而少量bin可能会导致大量信息丢失。

**论文提出的改进：**在MTF的原始表述中，Wang和Oates 2015提出将值的范围划分为多个相等大小的部分。由于在许多数据集中，值的分布趋于遵循长尾分布 (long tailed distributions)，因此，选择最佳bin的数量和宽度成为需要手动调整的挑战性步骤。较大的条带宽度会将大多数值聚合到最接近均值的条带中，而较小的条带宽度将导致极端条带中的稀疏性。为了克服对最佳参数设置的高度依赖，我们建议扩展bin选择和分配的原始公式，并改用执行非均匀bin的符号聚合近似 (SAX) 算法分配，使得频点频率的分布大致遵循高斯分布。因此，设置正确数量的仓bin并不是那么关键的任务，因为扩展bin选择和分配的原始公式后大大减少了稀疏风险，可以使用任意数量的仓，并且选择仓的数量主要取决于计算事务的计算要求。

## Recurrence Plot (RP)

递归图 (RP) (Kamphorst and Ruelle 1987) 将时间序列转换为递归矩阵，以揭示其中一些轨迹返回到先前访问过的状态。在论文中，我们遵循Souza, Silva和Batista 2014提出的RP的非二进制版本。RP编码包含以下转换：

$$RP_{i,j} = ||S_i - S_j||$$

Here, we set the sub-sequence length to 1 (we only consider the difference between single values), and defined  $K = l_2$ .

**论文提出的改进：**此转换是不可逆的，并且还会丢失时间序列的价值，因为RP仅保留值的差异。因此，我们建议进行以下转换：

$$RP = RP + \text{mean}(X) \text{ 此转换旨在捕获各个时间序列中均值附近的重复。}$$

## Gray-Scale (GS) Encoding 灰度图

2018年，Wen等人为了使用卷积神经网络对制造系统进行故障诊断恢复了灰度编码的使用。灰度图的转换首先包括将一维时间序列整形为大小为 $K_1$ 的 $K$ 个子序列的集合，其次是将值重新缩放为颜色编码值（例如8位整数）。在这项工作中，为了与其他编码保持一致，我们以 $K = K_1$ 为目标，并据此计算每个子系列开始之间的步幅。因此，GS编码对应于以下操作：

$$GS_{i,j} = \text{round} \left( P \cdot \frac{(x_{(i-1) \cdot s + j} - LB)}{UB - LB} \right), \quad i, j \in \llbracket 1, K \rrbracket,$$

**论文提出的改进：**在原始公式中，作者建议使用 $UB = \max(X)$ ， $LB = \min(X)$ 和 $P = 255$ ，以便将结果值解释为8位编码的灰度图像。在这项工作中，我们首先建议在整个训练数据集上定义**UB**和**LB**，并根据**GAF**编码选择这些范围。此外，由于大多数常见的深度学习实现均使用32位输入，因此我们建议评估此编码转换为整数的必要性。

## Spectrogram (SP) and Scalogram (SC)

为了补充上述编码，我们建议在研究中添加更传统的时频分析。由于时间序列的时空描述提供了二维表示，因此在文献中通常将这些表示直接解释为图像。在这项工作中，我们提出了两种流行的变换：**短时傅立叶变换（STFT）**产生的**频谱图**和**离散小波变换（DWT）**产生的**比例图**。

STFT和CWT都依赖于将时间序列与窗口函数进行卷积的原理。两种方法之间的区别在于，对于STFT，窗口是固定的，傅立叶变换是应用于卷积信号的；而对于CWT，窗口是基于小波函数的，其缩放比例使得能够探索不同频谱的特征。水平。因此，对于STFT，需要设置三个超参数：窗口类型，窗口大小和时间跨度。DWT需要先定义小波族，探索的尺度和时间跨度。

# Experiments

## DataSet

The use case on which we illustrate different ways to encode time series into images is relative to flight test helicopters vibration measurements.

数据集地址：<https://doi.org/10.3929/ethz-b-000415151>

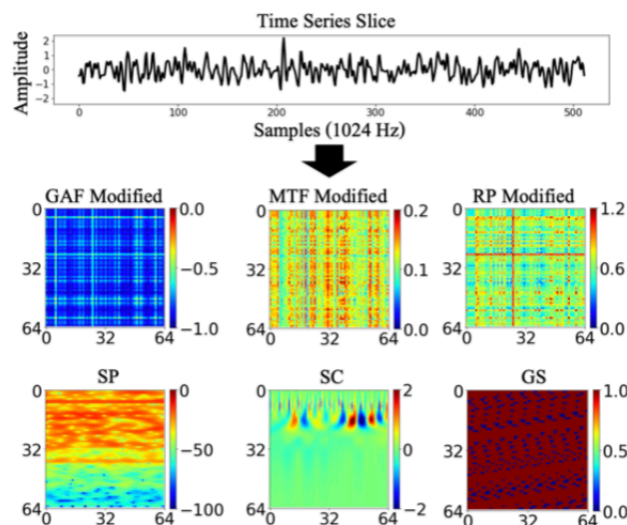


Figure 2: Encoding Examples.

重型仪器飞机（直升机或类似飞机）的飞行测试中的主要挑战是验证生成的数据数，因为要进行异常验证的信号数量比较庞大。手动验证需要太多时间和人力。此验证的自动化至关重要。

在这种情况下，将不同的加速度计放置在直升机的不同位置，且方向不同（纵向，垂直，横向），以测量直升机在所有运行条件下的振动水平。数据集由多个1D时间序列组成，这些时间序列从不同的飞行中以1024 Hz的恒定频率采集，分为1分钟的序列。标签（正常，异常）与每个航班相关联，并归属于航班的每个序列。

数据集分为两部分：

- 1、训练数据集：它仅包含正常飞行中的1677个序列；
- 2、验证数据集：它由正常和异常飞行的594个序列组成。

这里的挑战是在训练数据集中的正常状况监视数据上训练算法，该算法能够可靠地检测验证数据集中的异常（在训练过程中无法访问异常时间序列，并且没有任何关于数量和可能的异常类型）。

## Pre-processing and Hyperrameter Settings for the Time-series to Image Encodings

对于所有建议的编码，我们为120个大小为512的子系列的每一个生成一个图像。为了提高计算效率，我们在可能的情况下设置所有参数以获得大小为64×64的图像。根据定义，GAF MTF和RP编码的结果图像大小等于输入时间序列的图像大小。因此，我们使用average pooling缩小了生成的图像。

## Hyperparameters of the DL algorithms

For both the 1D and 2D CAE, we use the Adam optimizer with learning rate 0.001,  $\beta_1$  equal to 0.9 and mean squared error as the loss function. Additionally, both networks are trained for 10 epochs with batch size 200. For all convolutional and dense layers, a Leaky ReLU activation function is used with alpha equal to 0.3. The hyperparameters for the 1D network are obtained by following a grid search approach, while the hyperparameters for the 2D network are manually selected such that good reconstruction measures are observed for all types of encodings. Detailed information about the network architectures is given in Table 1.

## Evaluation Metrics

The performance metrics of all the selected setups are presented in Table 2.

Performance Measures					
Encodings	TPR	FPR	F1	AUC	Time[s]
No Encoding (1D)	0.61	<b>0.00</b>	0.76	0.80	-
GAF Original	0.08	0.10	0.14	0.49	24
GAF Modified	0.63	<b>0.00</b>	0.77	0.81	27
MTF Original	0.05	0.03	0.09	0.51	34
MTF Modified	0.74	<b>0.00</b>	0.85	0.87	264
RP Original	0.07	0.04	0.13	0.52	34
RP Modified	0.55	0.03	0.76	0.70	36
SP	<b>0.91</b>	0.41	0.78	0.75	4
SC	0.85	0.01	<b>0.91</b>	<b>0.92</b>	62
GS Original	0.11	0.06	0.18	0.52	1
GS - P=1	0.80	0.04	0.87	0.89	1
GS MinMax	0.65	<b>0.00</b>	0.79	0.82	1



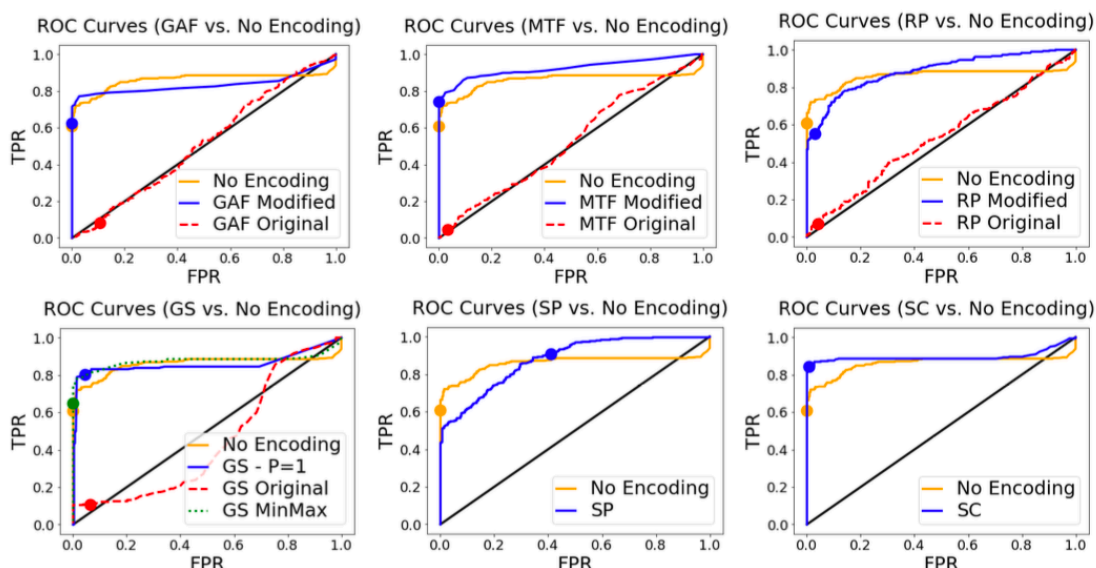


Figure 3: **ROC curves.** This figure illustrates the roc curves for the 2D-model subjected to all individual encoding well as for the 1D-model. The black diagonal line represents the random number generator baseline.

## Results

Figure 3和Table 2的分析中，我们可以得出结论，对于案例研究数据集和在论文建议的设置下，比例尺图SC是表现最佳的方法。它在F1和AUC指标上得分最高。

此外，与直接输入原始时间序列相比，对GAF，MTF，RP和GS编码的论文提出的修改均改善了异常检测算法的性能。大多数将时间序列转化为图像做的修改在于使此转换更好的为整个训练集服务，而不是像最初建议的那样将其切分输入的时间序列。这样做使得我们可以更好地捕获主类的特征行为。对于GS灰度图编码，找到正确的P值可能会对结果产生很大影响。先验地找到该参数是一个需要解决的问题。此外，不需要把灰度图的值转换为整数也优于基准。因此，在无人监督的任务中，选择合适的参数也是值得关注的部分。

根据F1度量标准比较，论文进行修正的所有时间序列编码表现均优于基准。请注意，F1度量是根据上面定义的建议阈值计算的。频谱图和递归图的AUC低于基线，这意味着阈值设置不同时，其优势也不确定。因此在无监督的检测任务中设置最佳阈值是一个开放性的研究问题。

在数据集级别上，所有方法均实现了非常低的FPR，这突出说明了主类学习得到了很好的学习。由于这些网络仅仅是基于正常情况的数据来设计和训练的，这一发现验证了他们对正常状况的表征能力。基于时间序列不同图像编码方式的主要区别主要在于它们能够更好地区分不同类型的异常，计算性能上的差异可能是由于某些编码更适合于某些类型的异常。

虽然时间序列的图像编码提供了性能方面的好处，部分也提供了可解释性方面的好处，但使用它们是以增加计算复杂性和所需内存为代价的。特别是SC和MTF，它们是性能最好的编码之一，需要相当多的时间来编码大量的时间序列。对于MTF，这一缺点仅适用于作为原始公式一部分的仓bin分配过程，而论文提出的扩展的SAX算法的克服了这一限制。

## Discussion

1、定义转换方式：在论文中，当采用原本未经改进的 编码方式进行异常检测时候，实验结果显示检测性能较差。定义每个输入的转换参数往往会使编码的图像更加统一。虽然在监督设置中学习非常特定的类鉴别特征和图像中的差异可能是有益的，但在无监督设置中，基于对单个（正常）类特征的学习，它是带有有偏见的。在执行转换时考虑到整个数据集的构成是实现所提出的编码的一个重要步骤。

2、阈值选择：只使用正常情况下的数据学习正确的阈值总是一项困难的任务。在我们的工作中，在残差上选择第99百分位数似乎提供了合理的结果，这符合阈值独立度量AUC。然而，我们强调，这一阈值仅在正常数据上学习。因此，很难说明这种选择对任何其他问题、任何其他类型的数据和异常都是有效的。3、选择时间序列的图像编码方式：尽管所有图像编码都提供了相当好的结果，但要事先说明哪种编码对特定用例来说是最好的还不容易。很明显，**每个编码将更适合于某些类型的时间序列，并将在各种类型的异常上表现不同**。4、CAE的体系结构：论文使用了一个相当简单的卷积神经网络，因为论文的主要目标是比较不同的编码和相同的体系结构被应用于所有的编码。探索其他体系结构来研究体系结构对性能的影响可能是未来的研究方向的。

5、图像编码的可解释性：与原始时间序列数据相比，使用时间序列进行图像编码的另一个优点是它们具有更直观的可解释性。

## Future Work

---

论文提出的框架被测试在特定类型的数据中。未来需要进行进一步的研究以评估在其他情况下，其他类型的数据和其他异常情况下的收益。

理想情况下，分析哪种编码最适合哪种数据类型和哪种异常类型的分析将对社区极为有益。基于这样的分析，可以在整体学习框架中组合编码以执行检测和诊断。