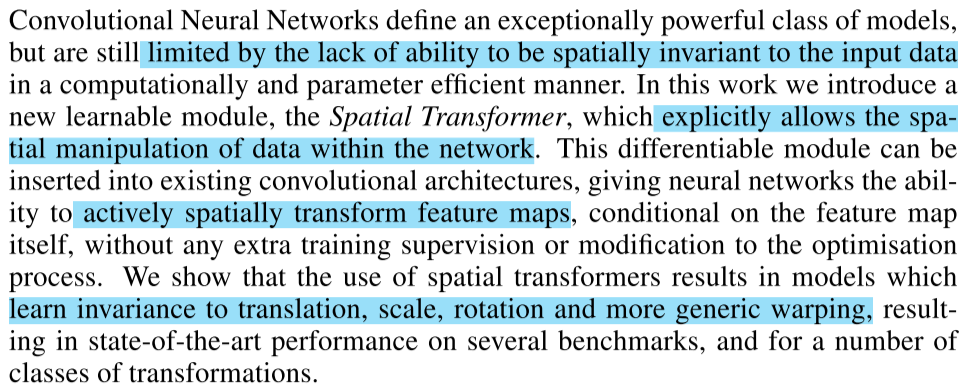
**STN阅读总结**

15331416 赵寒旭

**Abstract**



卷积神经网络定义了一个强大的分类模型，但是缺乏应对输入数据空间变换的能力。我们介绍了一种新的可习得的模型——空间变换，它明确允许了网络中数据的空间操作。这一可微的模型可以被插入目前的卷积结构，使神经网络能够对feature map进行基于其自身的空间变换而不需要对optimisation process进行任何额外的监督或修改。空间变换模型能够学习到平移，放缩，旋转和更多通用变换的不变性，提高网络的性能。

Q: spatial invariance具体含义是否指即使一个物体在空间上发生了变换，但是物体的属性仍然不变？比如原来的图像是数字1，即使经过平移旋转缩放等一系列空间变换，仍然保持为数字1并应该被网络识别。

试解：物理规律的空间平移不变性：如果在中国的实验室里做了某个实验，得到了某个物理结果，那么在美国的实验室里在完全相同的条件下做同一实验，必然会得到同样的结果。

对图像来说，空间不变性对应着图像处理的经典手段：平移、缩放和旋转，他们同属于空间变换，并可以通过坐标矩阵的仿射变换来实现。

仿射变换矩阵：

表示原图像像素点，表示仿射变换后图像的像素点。系数矩阵即为仿射变换系数，可以通过调整系数矩阵实现图像的平移缩放和旋转。

1. 平移

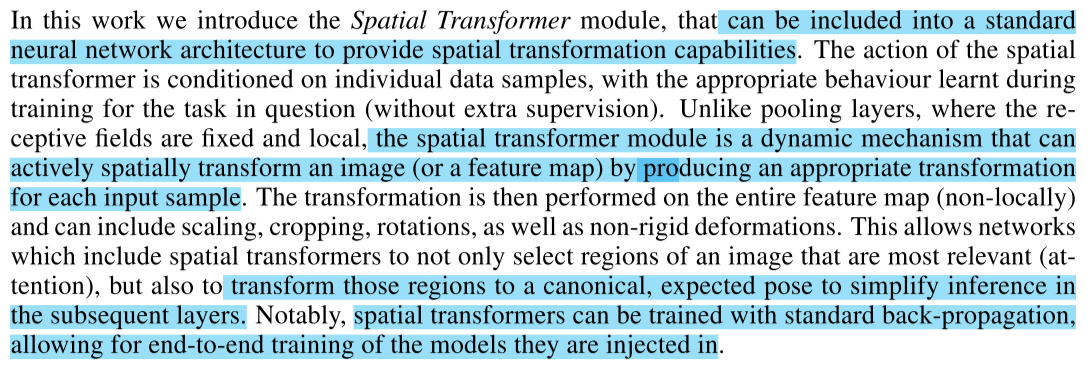
2. 缩放

3. 旋转

绕原点顺时针旋转度，坐标仿射矩阵：

需要做Normalization把坐标调整到使绕图像中心旋转。

**1. Introduction**

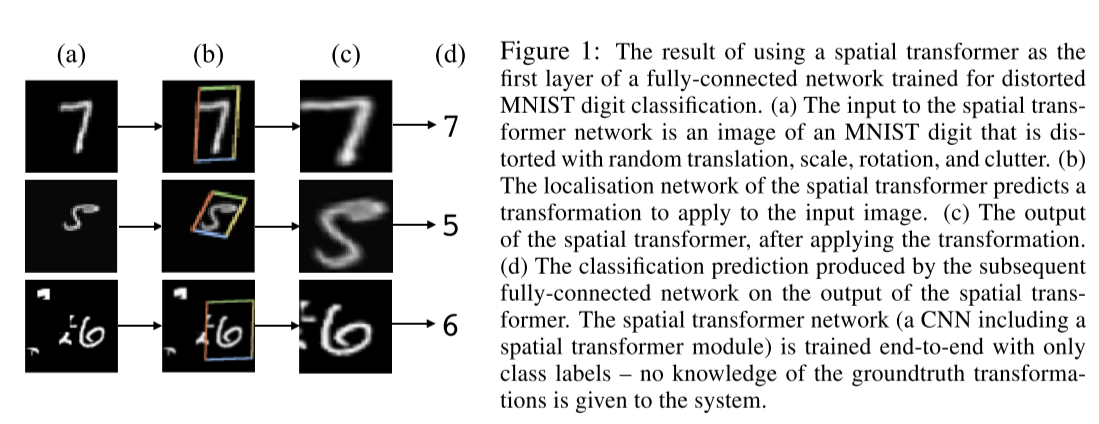


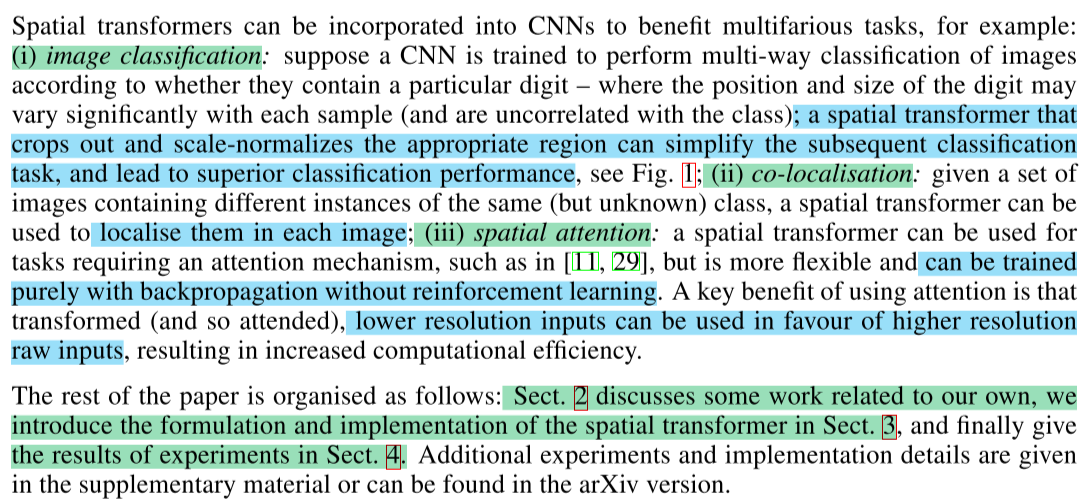
空间变换器模块可以被包含在标准神经网络结构中以提供空间变换能力。

空间变换器模块是一种动态机制，可以通过对每个输入样本产生合适的变换对一个图像（或者一个feature map）进行主动的空间变换。

包含空间变换的网络将不仅选择一个图片中最合适的区域，还可以把这些区域变换到一个标准的我们所期待的样子以简化后续层的推断。

空间变换可以被标准反向传播训练，允许包含这一模块的整个模型进行端到端的训练。





空间变换网络插入CNN中，对多种任务均有益：

（1）image classification

空间变换可以把输入剪切并缩放至合适的区域以简化后续的分类任务，提升分类性能。

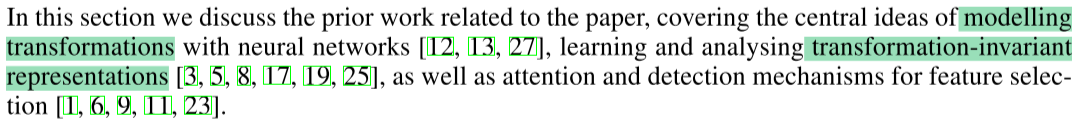
（2）co-localisation

给定一组包含一类物体不同实例的图像，空间变换器可以在每张图像中定位它们。

（3）spatial attention

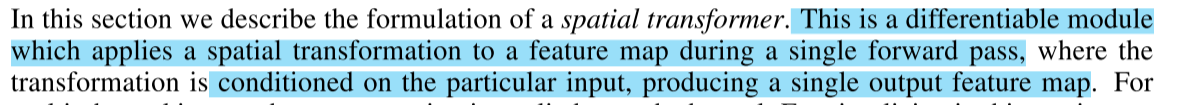
空间变换器可以被用于需要attention机制的任务中，它更加灵活并且可以只用反向传播进行训练。

**2. Related Work**

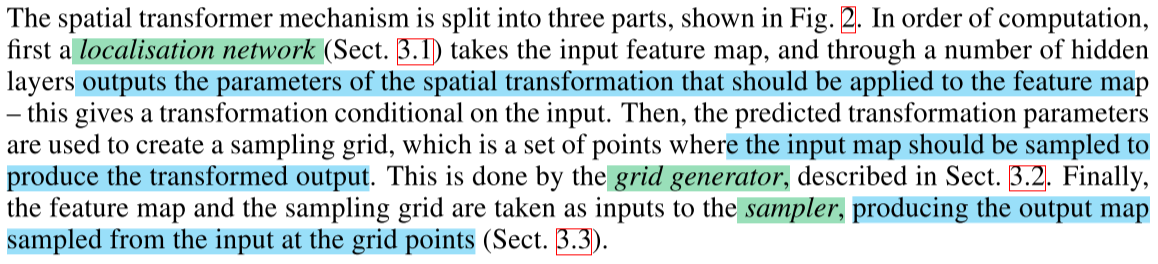


这一节是和本文有关的一些网络模型的讨论。

**3. Spatial Transformers**



这是一个在单个前向传递中把空间变换应用到feature map上的可微模型，其中变换基于特定输入产生单个输出feature map。

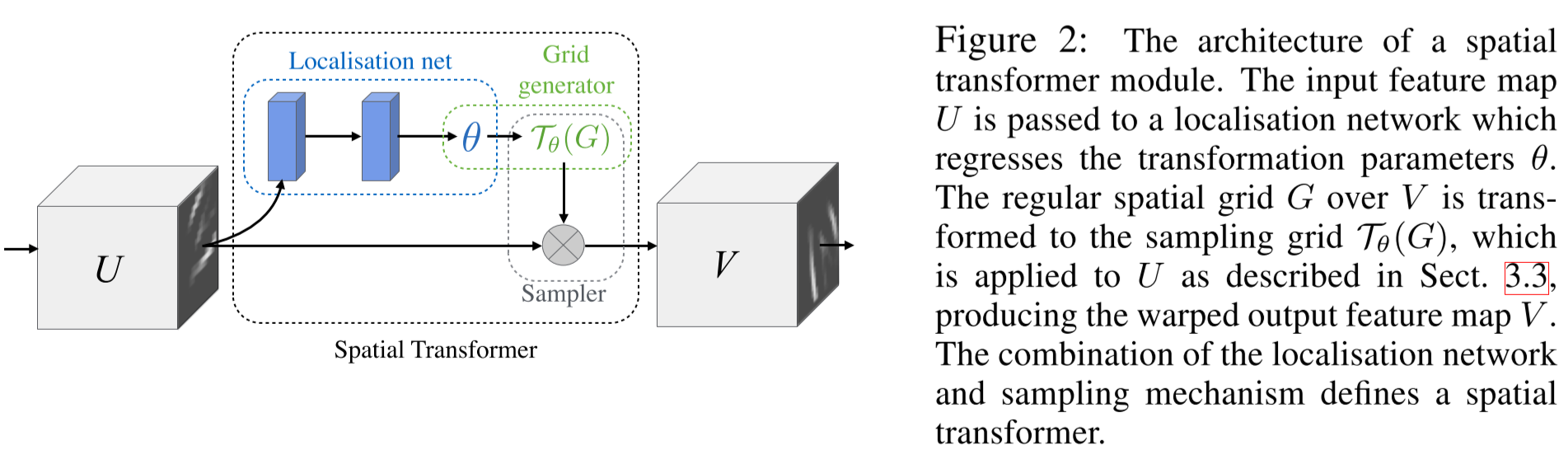


空间变换机制被分成三个部分：

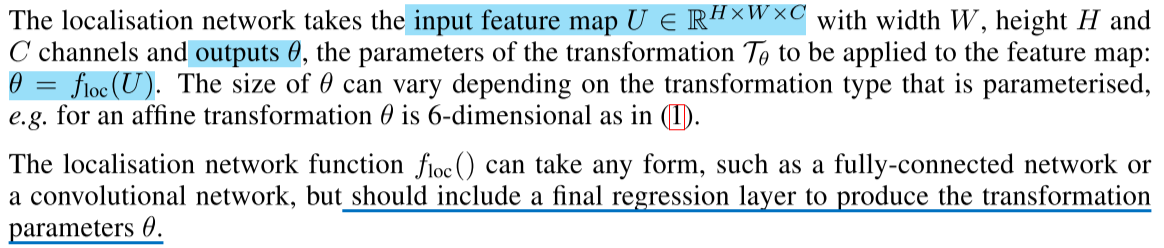
localization network: 输入feature map通过一系列隐藏层后输出要被应用于feature map的空间变换参数。

grid generator: 根据变换参数确定在输入feature map和输出feature map上的映射关系。

sampler: 结合输入的feature map和映射关系，获得变换后的输出。



**3.1 Localisation Network**



input: feature map （width W, height H, C channels）

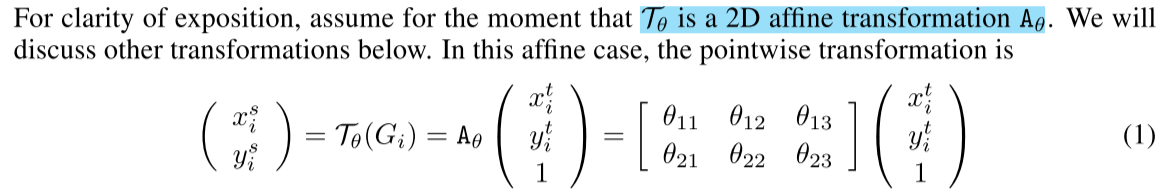
output:

通过一个子网络生成变换参数。

localisation网络函数可以是任何形式，如全连接网络或卷积网络，但最后一定有一个回归层用于生成变换参数。

的形式可以根据需要而变化，以2D仿射变换为例，就是一个2\*3的向量输出。

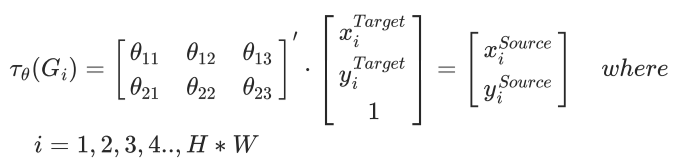
**3.2 Parameterised Sampling Grid**



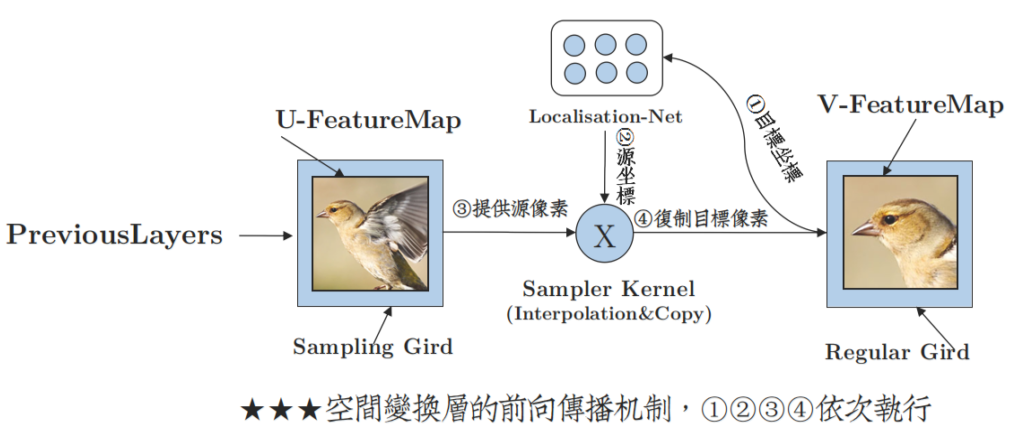
考虑逆向仿射变换，先根据仿射变换输出的大小，生成输出的坐标网格点，再对该坐标位置矩阵中的点进行仿射变换。

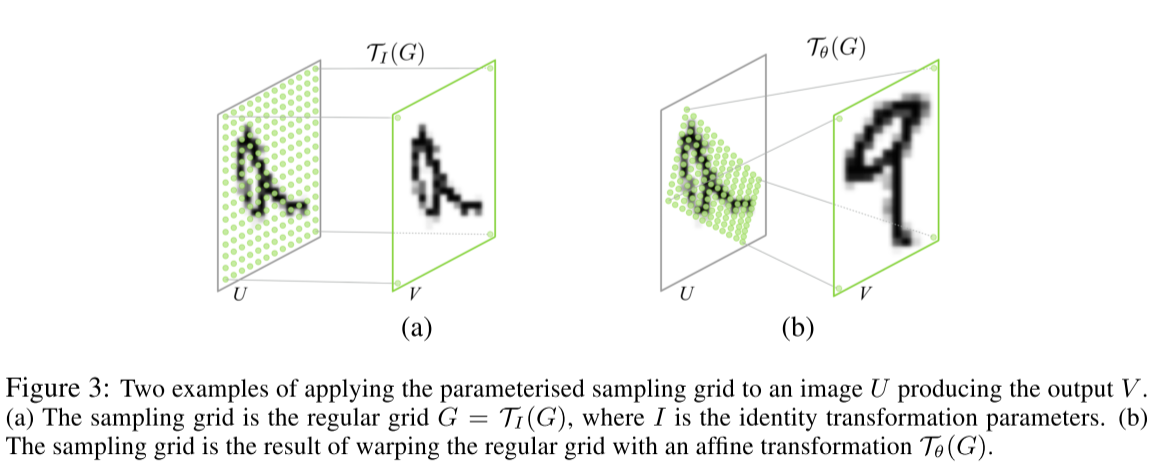
此时仿射系数为的逆矩阵，经仿射变换后可以得到V中的坐标点在U中的对应位置（可能非整数），再通过在输入图像中进行插值得到此坐标点的值。

得到U中坐标点的值后，则可将其复制到V中，得到仿射变换结果。



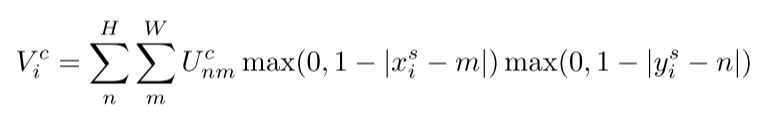
具体过程参考下图：





**3.3 Differentiable Image Sampling**

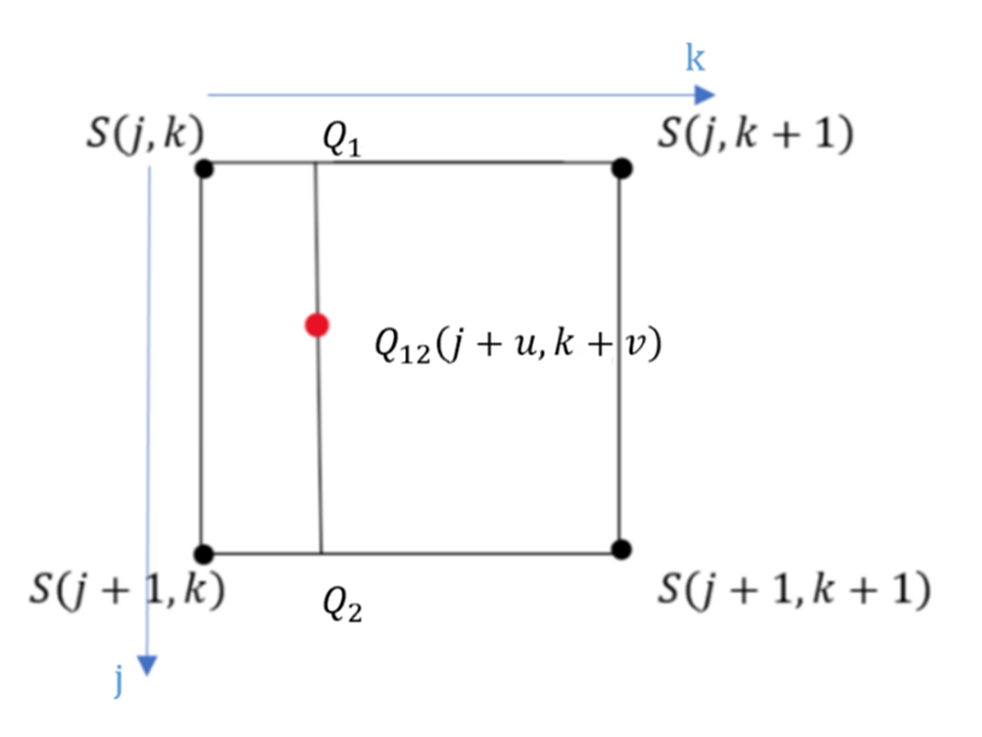
对双线性插值的情况有插值等式如下：

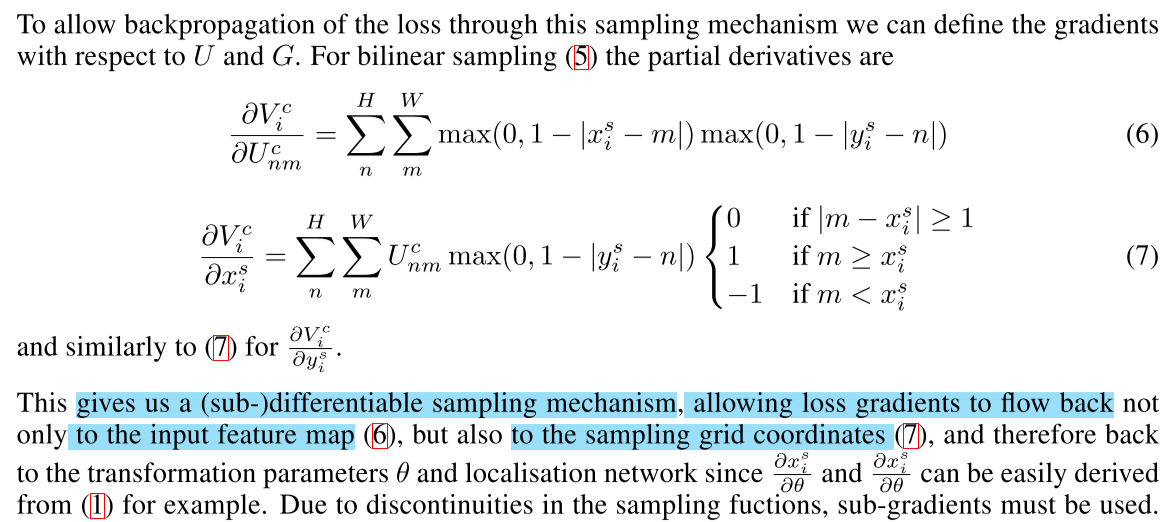


V中变换到U中，选取在U中的邻近点求和。

Q：此处不理解，按照双线性插值方法直接选择临近的四个点进行插值就可以了，而这个式子好像是循环选邻近点再求和。

双线性插值：

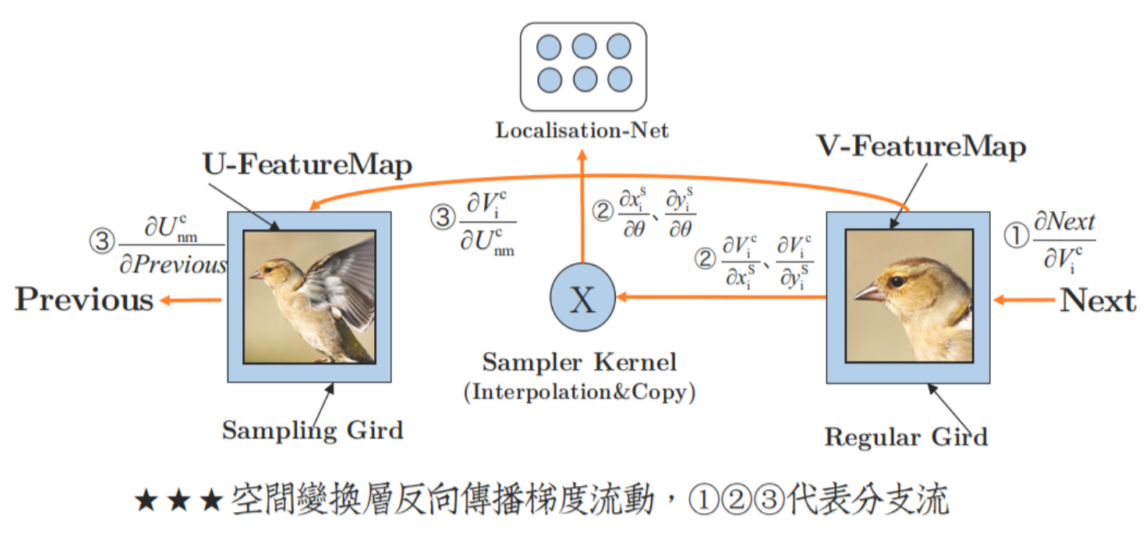




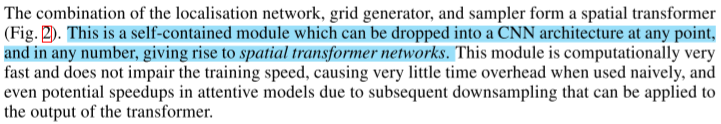
此处的采样核函数是不连续的，不能直接如下求导：

应该分两步，先对和求局部梯度、，后有：

梯度流动过程参考下图：



**3.4 Spatial Transformer Networks**

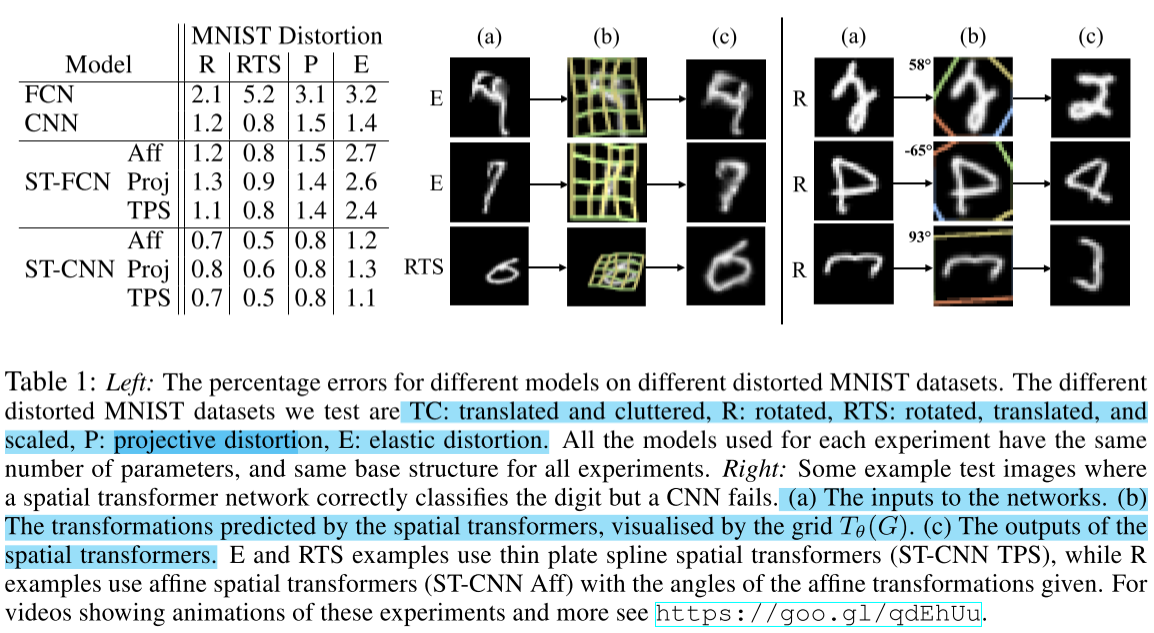


空间变换器是一个独立的模块，可以在任何时候放入CNN架构中，产生空间变换网络。这个模块的计算速度很快，不会损害训练速度。

**4. Experiments**

**4.1 Distorted MNIST**

本节中使用MNIST手写数据集作为测试平台，训练不同的神经网络模型来分类以方式失真的MNIST数据。使用空间转换的网络得到的结果优于其对应的基础网络。



上表展示了不同模型在不同畸变的MNIST数据集下的识别错误率结果。

**MNIST数据集畸变方式：**

R：旋转

RTS：旋转、缩放、平移

P：投射变换

E：弹性变形

**Model：**

使用两种基本的网络结构：FCN、CNN，先在无STN的情况下分别测试。

分别加入STN后，再进行实验。

STN采用的变换方法：

Aff：仿射变换

Proj：透射变换

TPS：薄板样条变换

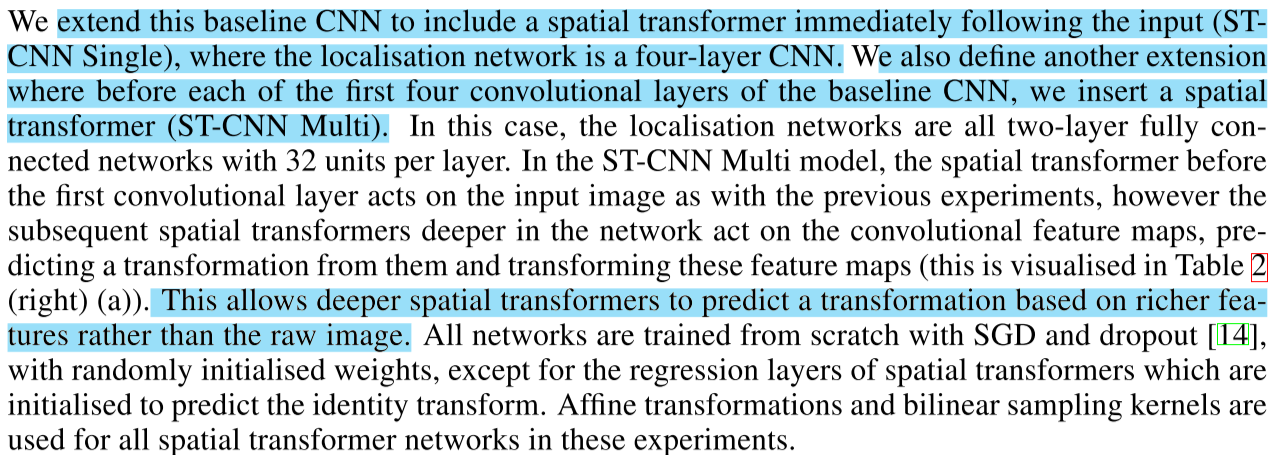
对数据集进行畸变的方式有多种，插入的空间变换网络也使用了多种变换方式。

可以看出，加入空间变换网络之后取得了更低的错误率。

右上图展示了STN对数字图像进行变换的结果，a)为原始输入数据，b)空间变换器预测的变换参数示意图，c)空间变换器输出的变换后结果。

**4.2 Street View House Numbers**

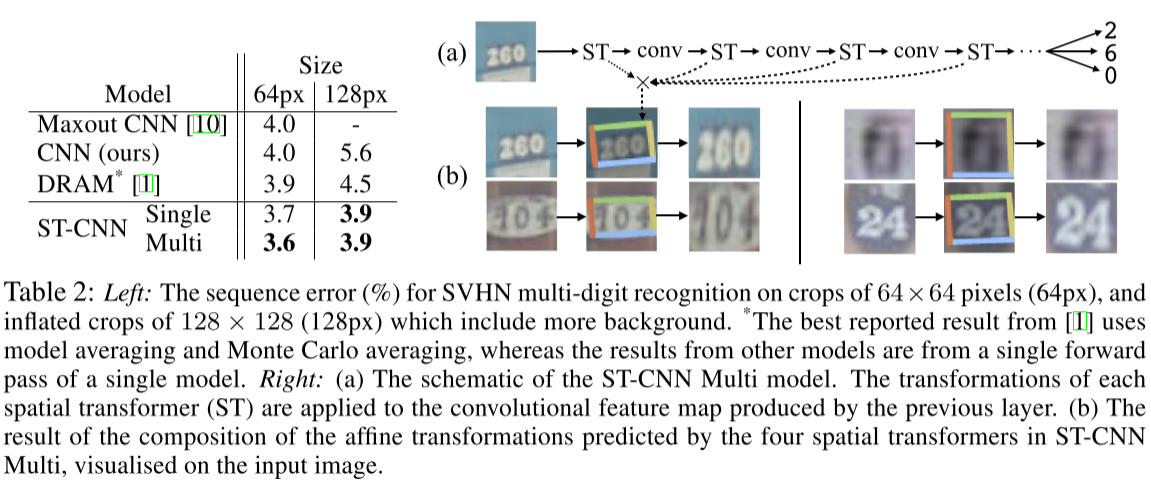
现在在一个具有挑战性的现实世界数据集Street View House Numbers（SVHN）上测试我们的空间变换网络。该数据集包含大约200k个房屋号码的真实世界图像，其任务是识别每个图像中的数字序列。 每幅图像中有1到5位数字，规模和空间布局变化很大。



ST-CNN Single: 将baseline CNN扩展为在输入后立即包含一个空间变换器。

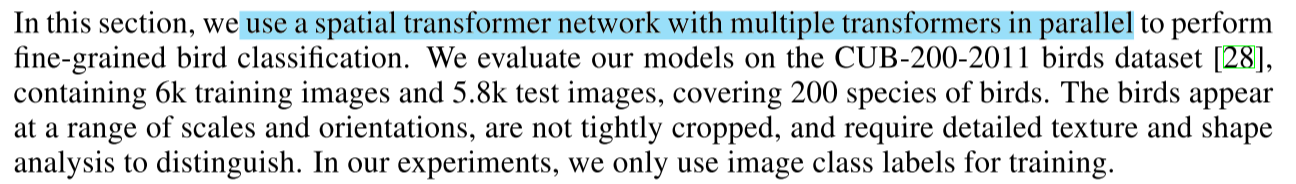
ST-CNN Multi: 在baseline CNN的前四个卷积层之前，各插入一个空间变换器。

在ST-CNN Multi中，第一卷积层之前的空间变换器如同前面的实验一样作用于输入图像，然而随后的网络中较深的空间变换器作用于卷积特征映射，预测它们的变换和变换这些特征图。这允许更深的空间变换器基于更丰富的特征而不是原始图像来预测变换。

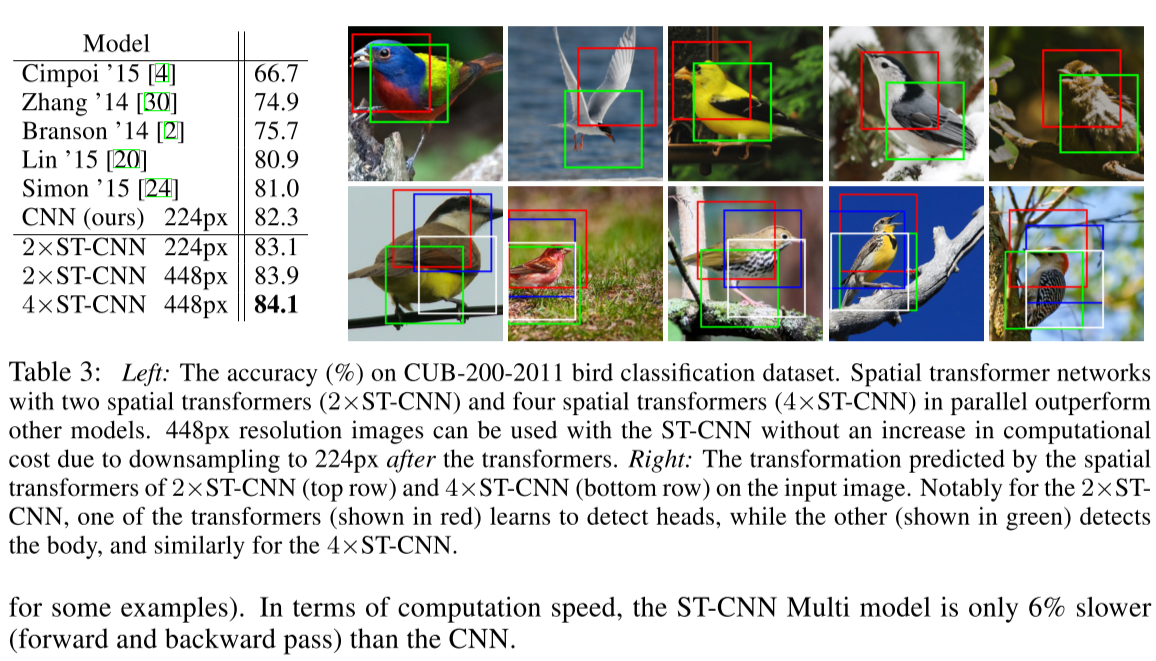


可见加入空间变换器的网络取得了更好的结果。

**4.3 Fine-Grained Classification**

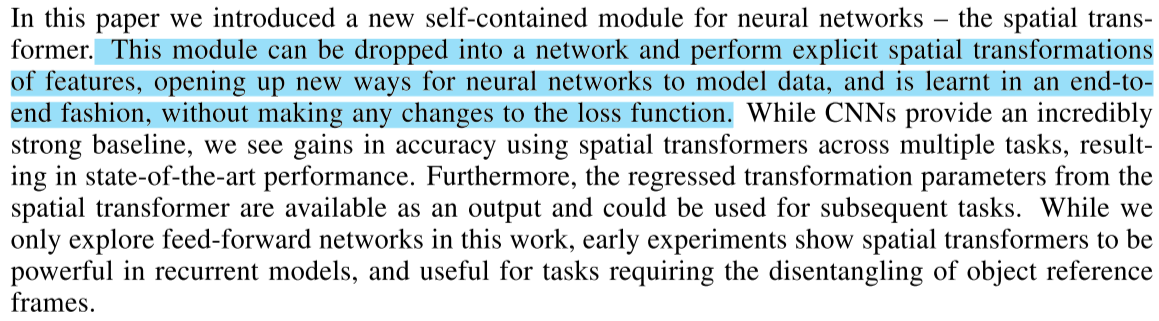


在本节中，我们使用具有多个变换器的的空间变换器网络并行执行细粒度鸟类分类。



具有4个并行空间变换器的空间变换网络有最高的准确率。

**5. Conclusion**



本文介绍了一种用于神经网络的新型独立模块：空间变换器。

这个模块可以放入网络中执行特征的显式空间转换，为神经网络建模数据开辟了新的途径，并以端到端的方式学习，而不对损失函数做任何改变。

使用跨多任务的空间变换器可以提高准确度，从而获得最先进的性能。

来自空间转换器的回归变换参数可用作输出，并可用于后续任务。