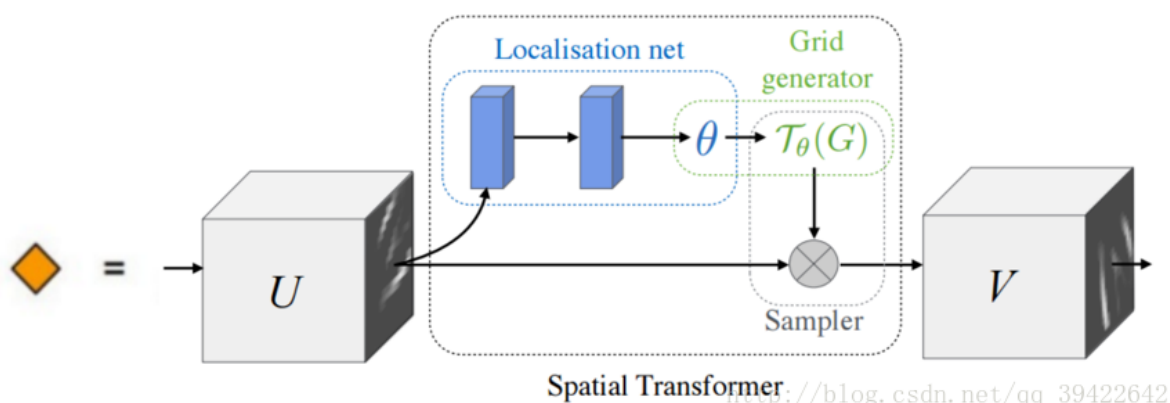


STN作用

普通的CNN有微小的平移不变性，STN网络在于用网络旋转缩放，使得模型具有平移不变性和旋转不变性。

STN架构



如图为模型主要架构，主要分为三个部分：

1. 参数预测：Localisation net
2. 坐标映射：Grid generator
3. 像素采集：Sampler

对应模型要解决的三个问题：

1. 寻找变换矩阵的参数
2. 设定变换后的图形/feature长和宽，找到原先坐标中的对应点（因此这里的输入为目标像素，输出为原像素）
3. 考虑2中可能计算出小数而实际坐标为整数，而四舍五入不可导，所以这里其实定义填充方式

参数预测

根据计算机图形学的知识，平移、缩放、拉伸等变换等线性变换都可以由一个仿射变换完成，无非是需要一个参数矩阵 W 。这个参数矩阵可以通过NN训练获得。

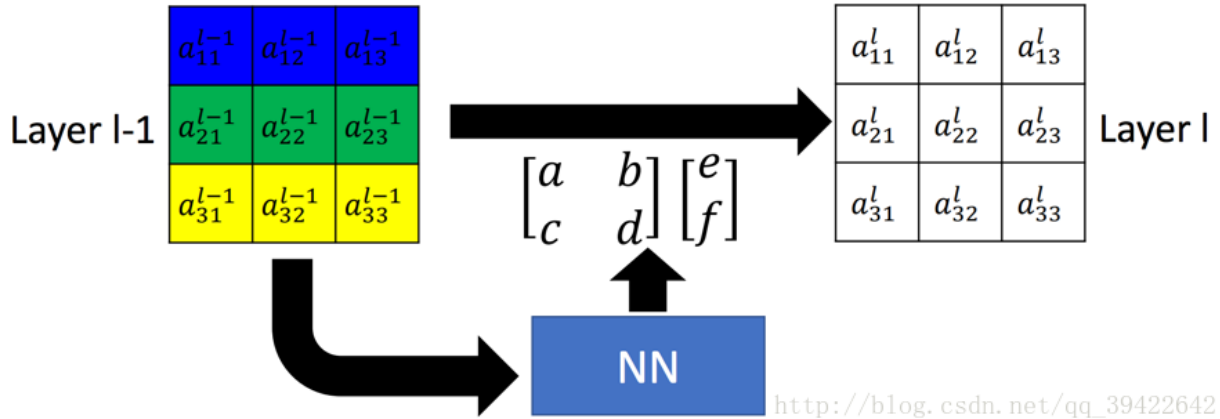
Grid generator坐标映射

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

6 parameters to describe the affine transformation

Index of layer l-1

Index of layer l



$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \Theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} \\ \Theta_{21} & \Theta_{22} & \Theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

Sampler

$$V_i = \sum_n \sum_m U_{nm} * k(x_i^s - m; \phi_x) * k(y_i^s - n; \phi_y)$$