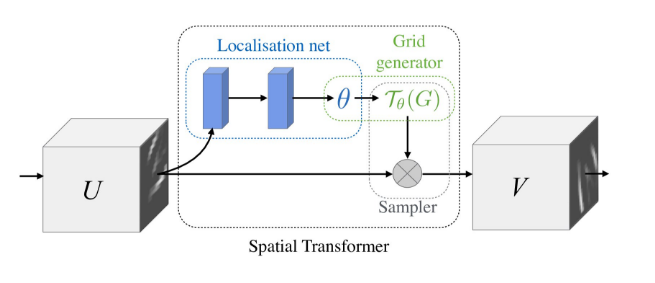
Spatial Transforms Networks（STN）

**STN结构：**

Affine Transformation涵盖平移、旋转和缩放三种变换操作，STN 能够实现对图像的平移、缩放、旋转以及裁剪，它主要有三部分结构组成：

Localisation net(作用:参数计算)、 Grid generator(作用:坐标映射) 、 Sampler(作用:像素采集)

①Localisation net: 是一个自定义的网络(可以是全连接或者卷积网，再加一个回归层)，它根据输入U，输出变化参数Ɵ，这个参数用来描述逆变换和映射U和V的坐标关系。

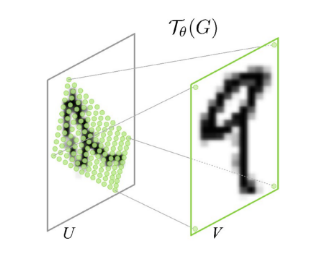
②Grid generator：根据V中的坐标点和变化参数Ɵ，计算出将要填充到V中的像素值在U中的坐标点。***注意：***这里是因为V的尺寸大小是自己先定义好的，当然可以得到V的所有坐标点，而填充V中每个坐标点的像素值的时候，要从U中去取，所以根据V中一个坐标点和变化参数Ɵ进行运算，得到U中的一个坐标。

③Sampler：要做的是填充V，根据Grid generator得到的U中一系列坐标和原图U（因为像素值要从U中取）来填充。如果计算出来的坐标值为小数，可以用另外的方法来填充，比如双线性插值。

**具体变换：**

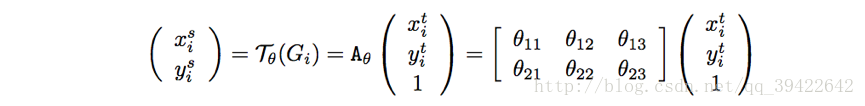
1. 确定坐标

根据变换参数Ɵ，在原样本上采样，拿到对应的像素点。通俗点说，就是输出的图片(i,j)的位置上，要对应输入图片的哪个位置。



如图所示旋转缩放操作，我们把像素点看成是坐标中的一个小方格，输入的图片U可以是一张图片或feature map，经过变换，(Ɵ是Localisation net生成的参数)，生成了图片V，它的像素相当于被贴在了图片的固定位置上，用表示，像素点的位置可以表示为，这就是我们要确定的坐标。

2. 仿射变换关系



是输出的目标图片的坐标,，

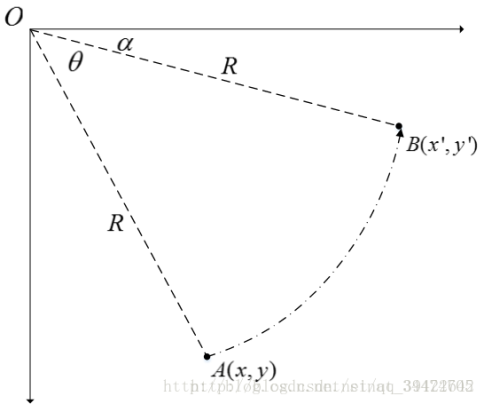
***注意***：坐标映射关系是

从目标图片→原图片

也就是说，坐标的映射关系是从目标图片映射到输入图片上的。

原因：坐标映射的作用。其实是让目标图片在原图片上采样，每次从原图片的不同坐标上采集像素到目标图片上，而且要把目标图片贴满，每次目标图片的坐标都要遍历一遍，是固定的，而采集的原图片的坐标是不固定的，因此用这样的映射。

**补充：**仿射变换中的旋转操作



由点A旋转θ度角，到达点B.得到

*x′= Rsinα*

*y′=Rsinα*

由A点得

*x=Rcos(α+θ)*

*y=Rsin(α+θ)*

联立两式得

*x=x′ cosθ−y′sinθ*

*y=y′cosθ+x′sinθ*

