

Chapter 1

Spatial Transformer Networks

The use of spatial transformers results in models which learn invariance to translation, scale, rotation and more generic warping, resulting in state-of-the-art performance on several benchmarks, and for a number of classes of transformations.

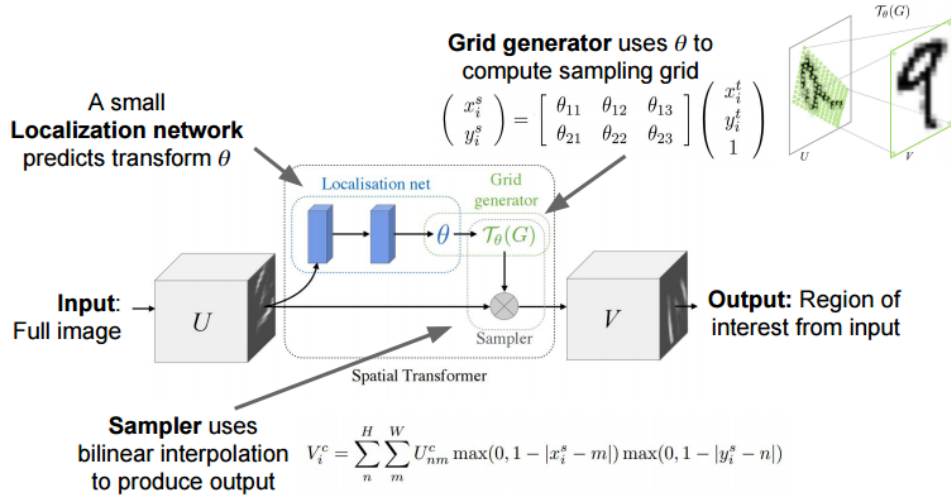


FIGURE 1.1: Spatial Transformer Networks

We would like a function that given a continuous box coordinates cropped the input image. The problem is that it is not possible to crop in a continuous way because we are constraint, at least, by pixels.

The idea is to have a function that maps output pixels coord to input pixels coord. This function is deferentiable w.r.t the affine parameters θ .

With this function, the network will attend to input regions by predicting θ .

They put the info explained on top row in a self contained module which they call "Spatial Transformer" which is divided in three parts:

- Localisation net: Outputs the affine parameters θ .

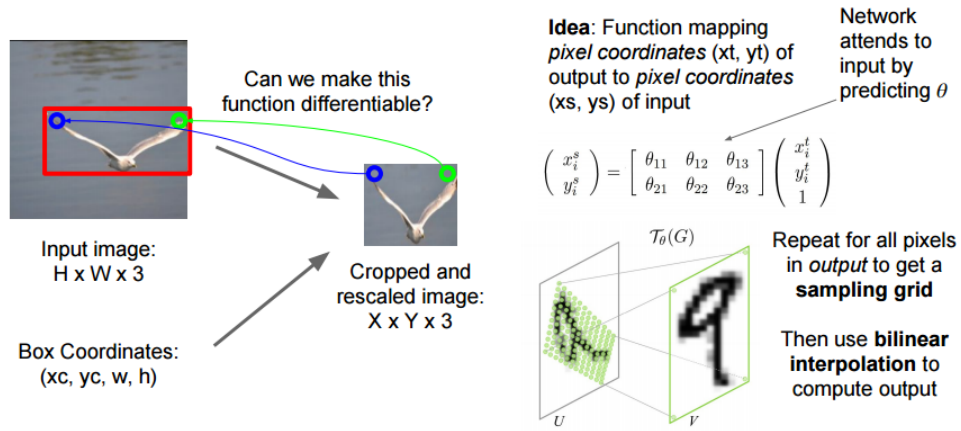


FIGURE 1.2: continuous coordinates

- Grid Generator: Use θ to compute the sampling grid
- Sampler: With bilinear interpolation produces outputs.

Notice that all this modules are continuous and differentiable.

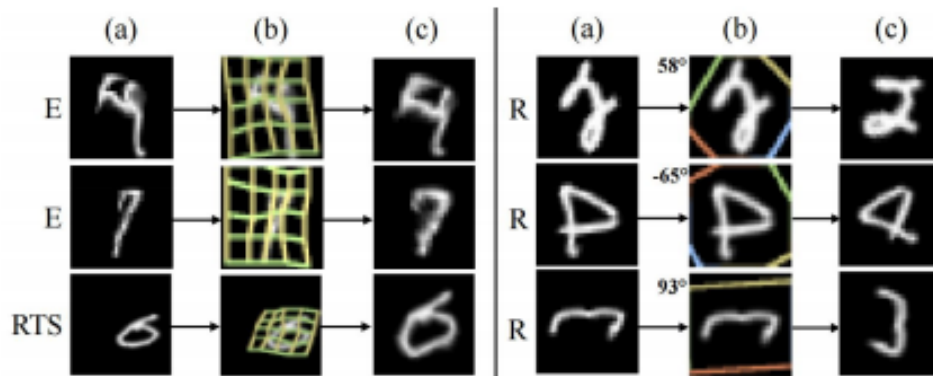


FIGURE 1.3: Deformation examples