Spatial Transformer Networks

Mario Teixeira Parente (m.parente@fz-juelich.de)

ISIS LENS Machine Learning School
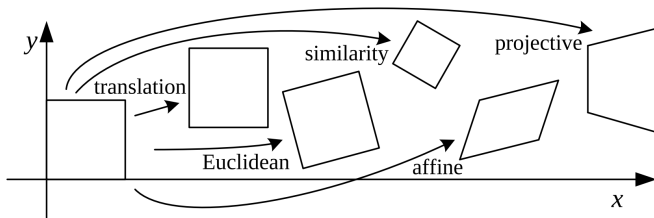
February 19, 2021

Introduction

- A **Spatial Transformer Network** (STN) aims to make a network *spatially invariant* to its input data.
    - $\implies$ More accurate object classification.
    - $\implies$ Localization of objects in an image and sub-classification.

- So far, CNNs are only somewhat invariant in translation through the Max-Pooling layer.

# $2D$ **geometrical transformations**

1. Translation
2. Euclidean/Rigid transformation (Rotation + Translation)
3. Similarity (scaled rotation + translation)
4. Affine transformation
5. Projection

Spatial Transformer Module

Spatial Transformer

(a)

(b)
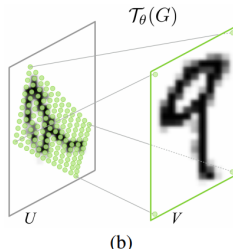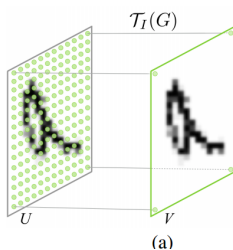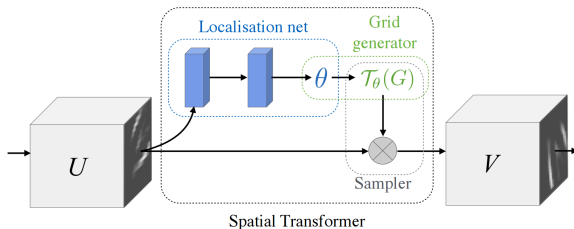
Results

		MNIST Distortion			
Model		R	RTS	P	E
FCN		2.1	5.2	3.1	3.2
CNN		1.2	0.8	1.5	1.4
ST-FCN	Aff	1.2	0.8	1.5	2.7
	Proj	1.3	0.9	1.4	2.6
	TPS	1.1	0.8	1.4	2.4
ST-CNN	Aff	0.7	0.5	0.8	1.2
	Proj	0.8	0.6	0.8	1.3
	TPS	0.7	0.5	0.8	1.1
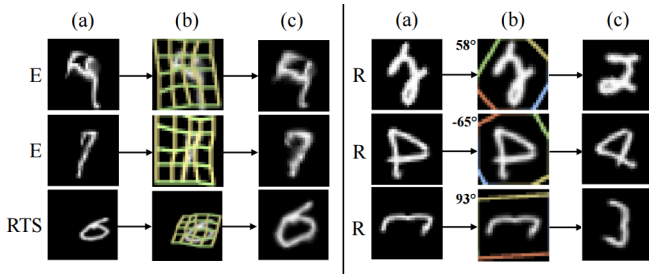
Results – II

Model	
Cimpoi '15 [5]	66.7
Zhang '14 [40]	74.9
Branson '14 [3]	75.7
Lin '15 [23]	80.9
Simon '15 [30]	81.0
CNN (ours) 224px	82.3
2×ST-CNN 224px	83.1
2×ST-CNN 448px	83.9
4×ST-CNN 448px	**84.1**
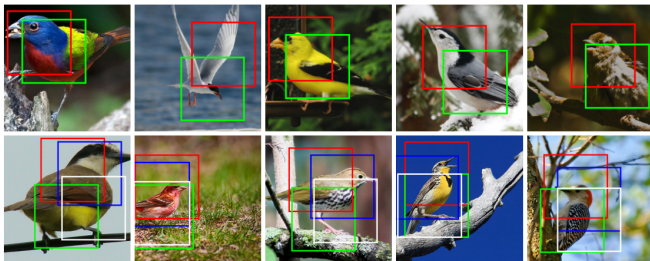
Thank you for your attention!

The **exposition** (also images) followed a
TUM wiki documentation (`link`).

A **Jupyter notebook** for learning
can be downloaded at `pytorch.org` (`link`).