Inverse Compositional Spatial Transformer Networks
Chen-Hsuan Lin, Simon Lucey
Carnegie Mellon University

Overview
Typical CNNs tolerate spatial variations within data implicitly through
1. Data augmentation (generate geometric perturbations)
2. Spatial pooling (abstracts semantics, but destroys spatial details)
We propose IC-STN to resolve spatial variations explicitly via recurrent transformations

Spatial Transformer Networks (STN)
learns to resolve spatial variations explicitly by predicting the geometric transformation \( p \) on the input image \( I_{in} \)

- Boundary effect
  - Geometry is not preserved throughout the network

- Single prediction
  - Appearance decorrelates with larger geometric displacements

The Lucas-Kanade (LK) Algorithm
solves for alignment by iteratively predicting updates to the warp \( p \) on image \( I \) to match the template \( T \)

- Original form
  \[
  \min_{\Delta p} \left\| I(p + \Delta p) - T(0) \right\|^2
  \]

- Inverse Compositional (IC)
  \[
  \min_{\Delta p} \left\| I(p) + \left( \frac{\partial I(p)}{\partial p} \right) \Delta p - T(0) \right\|^2
  \]

Iteration-specific linear models (dependent on warp state \( p \))
Static linear model (independent on warp state \( p \))

IC-STN
- preserves geometry and original input image
- utilizes a single geometric predictor
- learns recurrent spatial transformations

Experiments (perturbed MNIST classification)

Experiments (traffic sign classification)

Discussions
- Theoretical connection: STN \( \leftrightarrow \) IC-LK
- Tolerating data spatial variations needs huge increase of model capacity
- Alignment is more efficient predicting small geometric updates iteratively

Check out our paper and code for more details and discussions!

Code available!
https://github.com/ericlin79119/IC-STN

Mean appearance

Variance heatmap

Classification error

Initial