SPATIAL TRANSFORMER NETWORKS

CS574: Computer Vision using Machine Learning by Prof. Arijit Sur

---

130101018 Desh Raj
130101041 Samyak Kumbhalwar
130101062 Sumeet Ranka
130101072 Siddharth Kumar
130101088 Akashdeep Goswami

September 13, 2017

Indian Institute of Technology Guwahati

Convolutional Neural Networks in Vision

CONVOLUTIONAL NEURAL NETWORKS IN VISION

CNNs have been extensively used in computer vision applications such as:

- Object detection
- Image classification
- Semantic segmentation

Have outperformed state-of-the-art learning methods for these tasks.

CONVOLUTIONAL NEURAL NETWORKS IN VISION

Multiple convolutional layers with local max-pooling layers allows some translational invariance.

However, **what should be done for highly distorted inputs?**

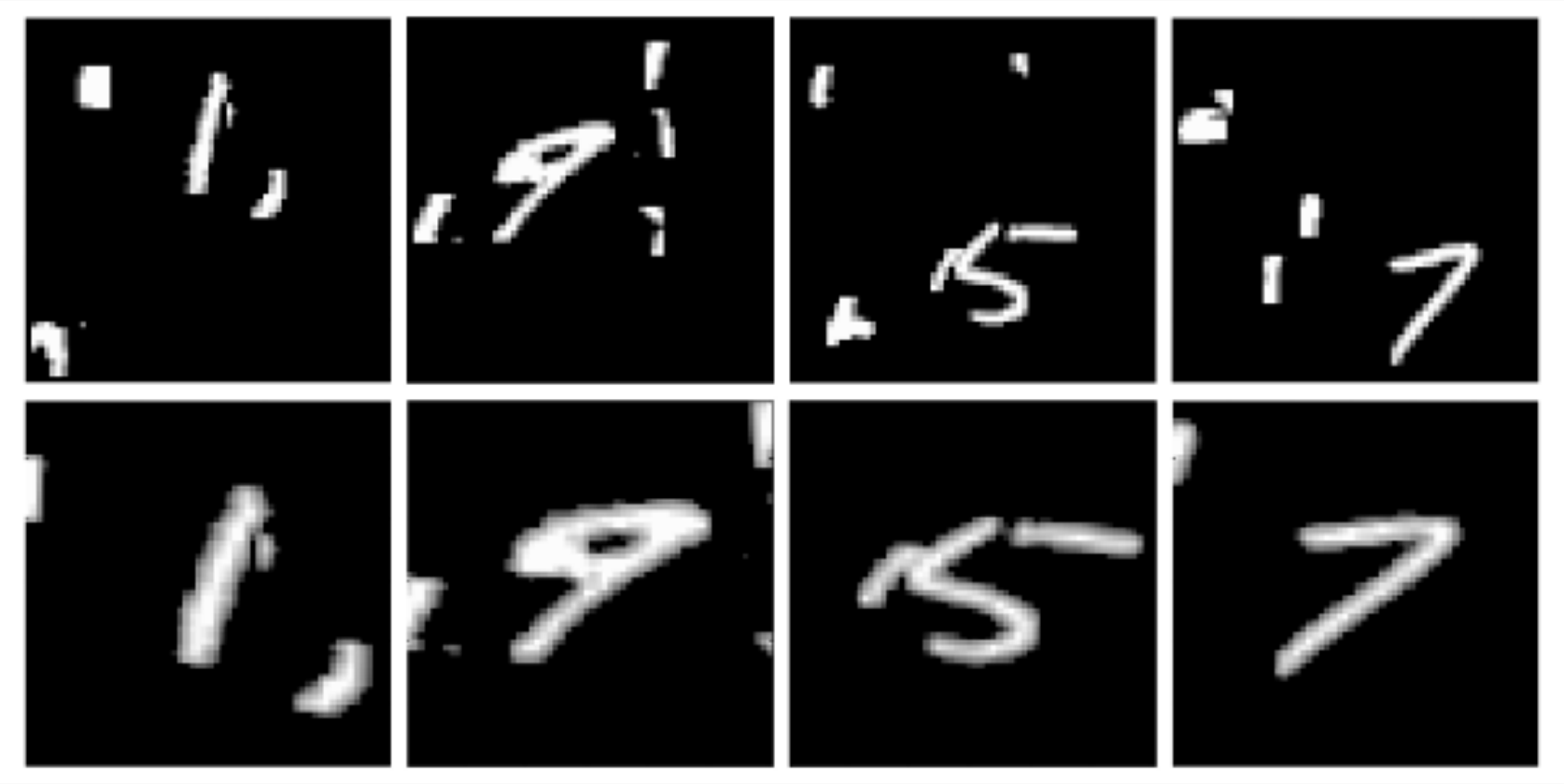

Figure: Distorted MNIST handwritten character data set.
(http://www.vlfeat.org/matconvnet/spatial-transformer/)

CONVOLUTIONAL NEURAL NETWORKS IN VISION

- Pooling is simplistic

CONVOLUTIONAL NEURAL NETWORKS IN VISION

- Pooling is simplistic
- Small invariances per pooling layer

CONVOLUTIONAL NEURAL NETWORKS IN VISION

- Pooling is simplistic
- Small invariances per pooling layer
- Limited spatial transformation

CONVOLUTIONAL NEURAL NETWORKS IN VISION

- Pooling is simplistic
- Small invariances per pooling layer
- Limited spatial transformation
- Limited spatial invariance provided by CNN!

Conditional Spatial Warping

CONDITIONAL SPATIAL WARPING

Conditional on input feature map, spatially warp the data.

- Transforms data into a space expected by subsequent layers

CONDITIONAL SPATIAL WARPING

Conditional on input feature map, spatially warp the data.

- Transforms data into a space expected by subsequent layers
- Select regions of attention

CONDITIONAL SPATIAL WARPING

Conditional on input feature map, spatially warp the data.

- Transforms data into a space expected by subsequent layers
- Select regions of attention
- Invariant to more classes of transforms

CONDITIONAL SPATIAL WARPING

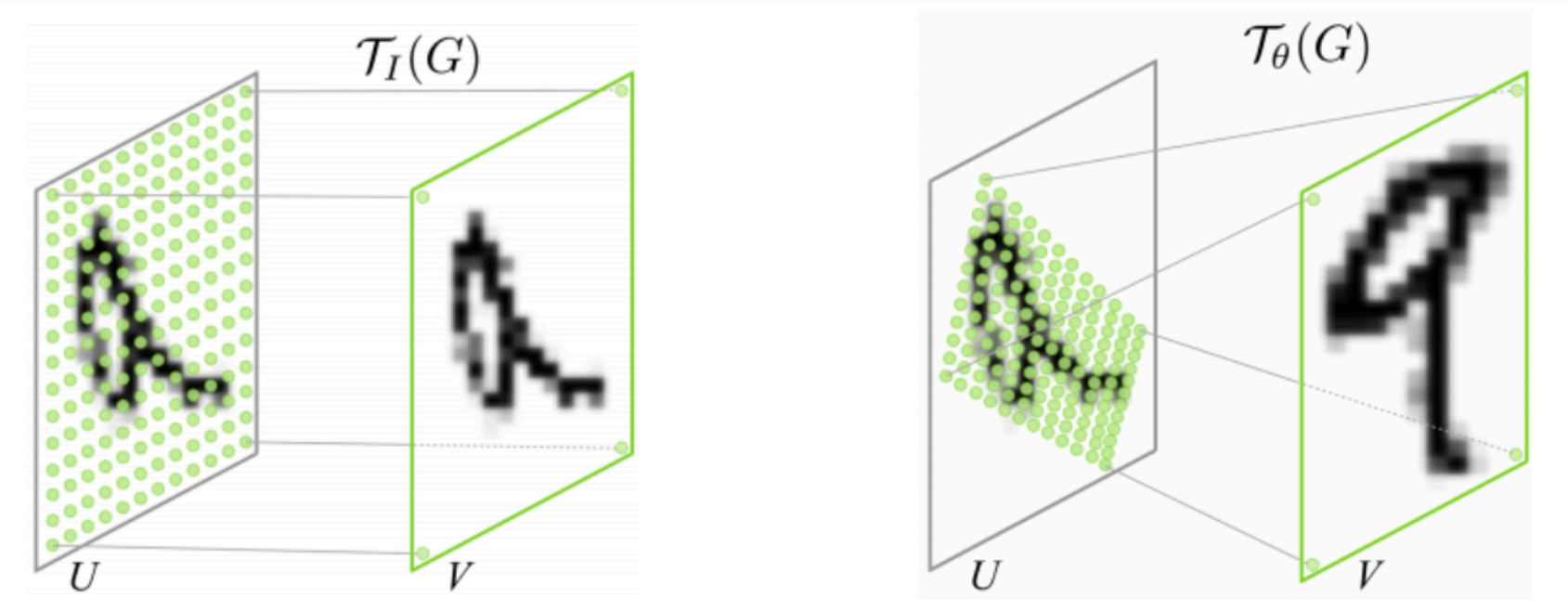

Figure: Transforming the attention grid. (Jaderberg, Max, et al. "Spatial Transformer Networks." arXiv preprint arXiv:1506.02025)

Spatial Transformer Network

SPATIAL TRANSFORMER NETWORK

- Introduced by Jaderberg et al. (researchers at Google DeepMind) at NIPS 2015.

SPATIAL TRANSFORMER NETWORK

- Introduced by Jaderberg et al. (researchers at Google DeepMind) at NIPS 2015.
- A completely independent module that can be plugged into any existing network.

SPATIAL TRANSFORMER NETWORK

- Introduced by Jaderberg et al. (researchers at Google DeepMind) at NIPS 2015.
- A completely independent module that can be plugged into any existing network.
- Various classes of transforms may be applied to input image to feed into further layers.

SPATIAL TRANSFORMER NETWORK

- Introduced by Jaderberg et al. (researchers at Google DeepMind) at NIPS 2015.
- A completely independent module that can be plugged into any existing network.
- Various classes of transforms may be applied to input image to feed into further layers.
- Learns using backpropagation, without explicit supervision.

SPATIAL TRANSFORMER NETWORK

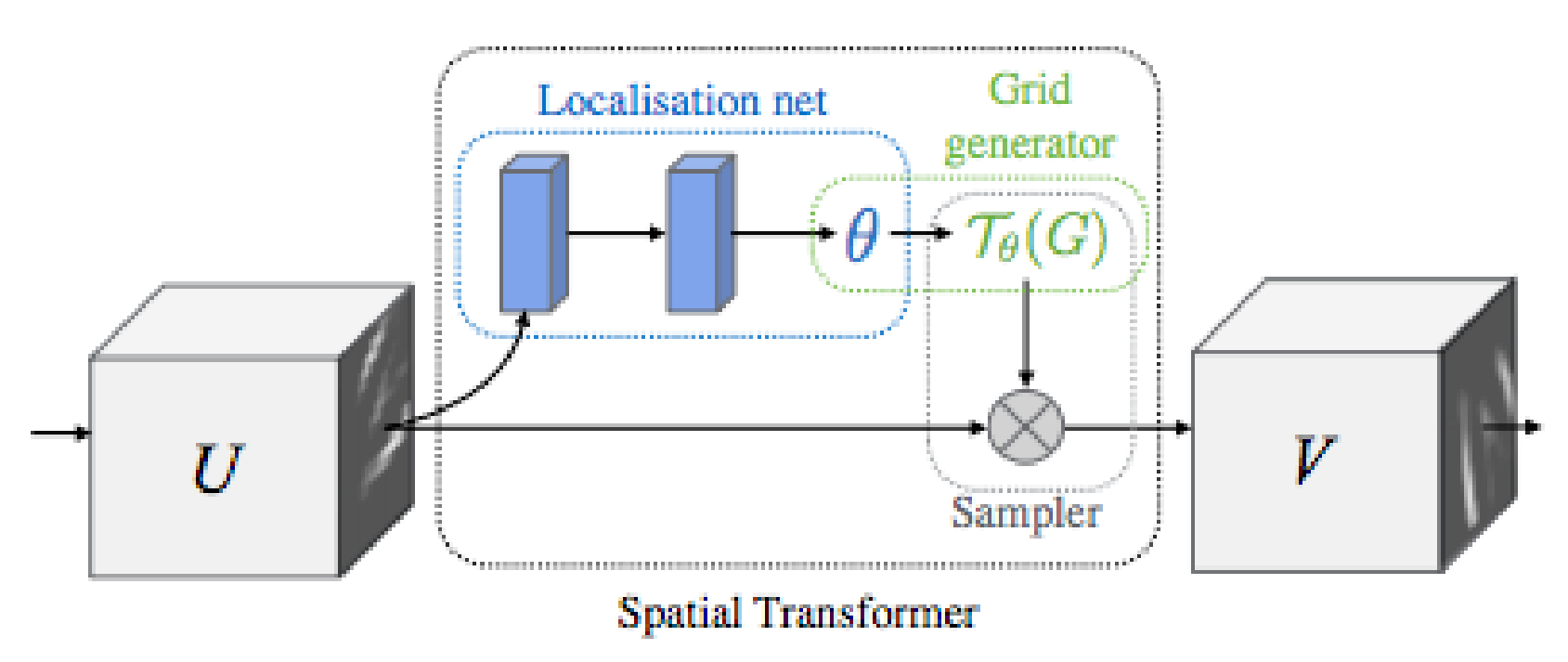

Figure: Architecture of an STN (Jaderberg, Max, et al. "Spatial Transformer Networks." arXiv preprint arXiv:1506.02025)

SPATIAL TRANSFORMER NETWORK

Localisation network Learns transformation parameter θ using the input feature map U.

SPATIAL TRANSFORMER NETWORK

Localisation network Learns transformation parameter θ using the input feature map U.

Grid generator Produces a sampling grid from the regular grid G and the transformation matrix τ_θ.

SPATIAL TRANSFORMER NETWORK

Localisation network Learns transformation parameter θ using the input feature map U.

Grid generator Produces a sampling grid from the regular grid G and the transformation matrix τ_θ.

Sampler Produces the output image from the input image and sampling grid.

SPATIAL TRANSFORMER NETWORK

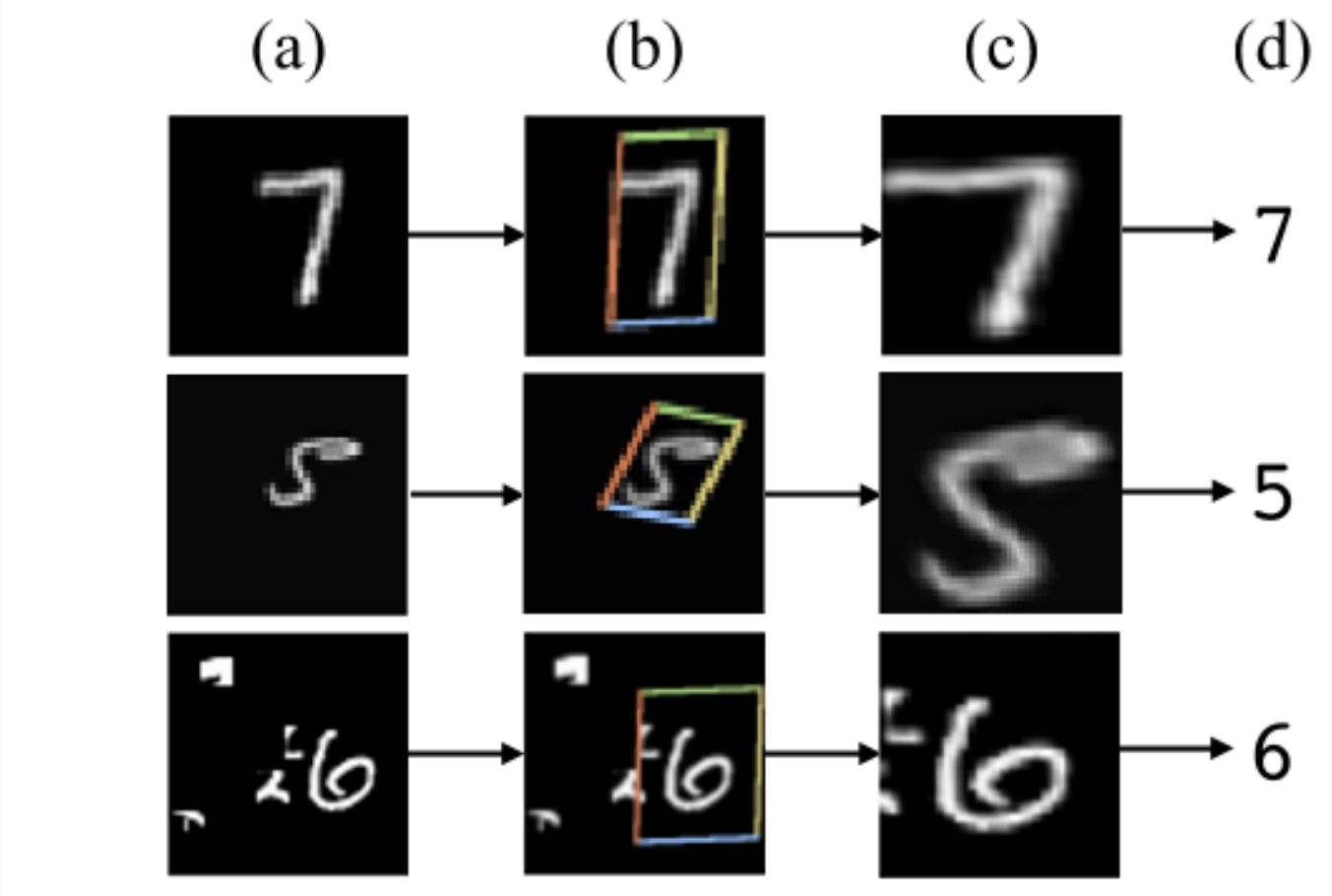

Figure: MNIST handwritten character recognition using a CNN-STN model (Jaderberg, Max, et al. "Spatial Transformer Networks." arXiv preprint arXiv:1506.02025)

Variations to STN: a Recurrent Model

VARIATIONS TO STN: A RECURRENT MODEL

- Proposed by Sønderby et al. (University of Copenhagen, Denmark).

VARIATIONS TO STN: A RECURRENT MODEL

- Proposed by Sønderby et al. (University of Copenhagen, Denmark).
- Uses a simple RNN in the localisation network instead of CNN as proposed in the first paper.

VARIATIONS TO STN: A RECURRENT MODEL

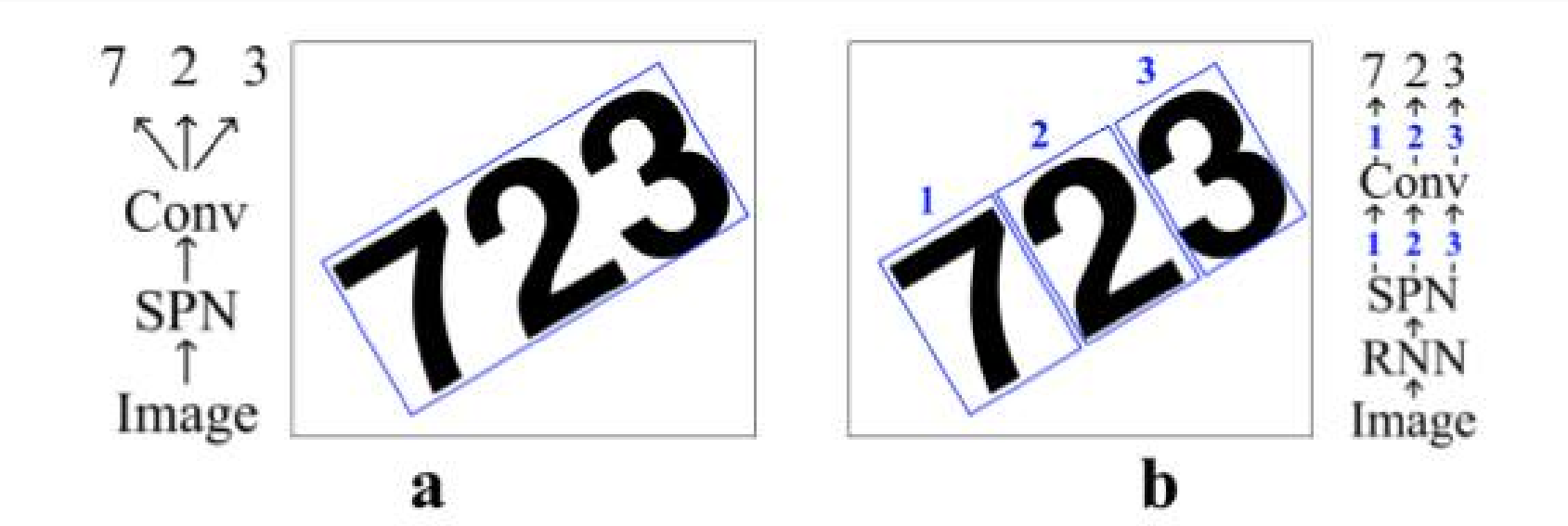

Figure: Architecture of the RNN-STN.

Proposed Improvements

PROPOSED IMPROVEMENTS

- Extend recurrence to the entire STN module instead of just the localisation network.

PROPOSED IMPROVEMENTS

- Extend recurrence to the entire STN module instead of just the localisation network.
- Allows multiple glimpses over the input image.

PROPOSED IMPROVEMENTS

- Extend recurrence to the entire STN module instead of just the localisation network.
- Allows multiple glimpses over the input image.
- Current glimpse is used as input to the RNN for the next iteration.

PROPOSED IMPROVEMENTS

- Extend recurrence to the entire STN module instead of just the localisation network.
- Allows multiple glimpses over the input image.
- Current glimpse is used as input to the RNN for the next iteration.
- Expected to improve classification accuracy, especially with images containing multiple digits.

Thank You