dortmund university

SFB 876 Providing Information by Resource-Constrained Data Analysis

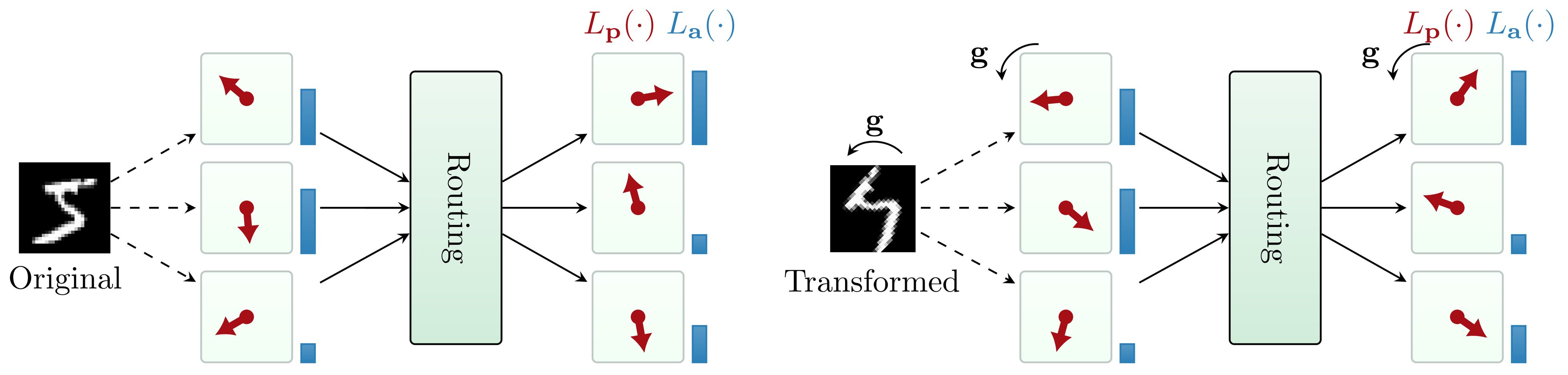
Group Equivariant Capsule Networks

Jan Eric Lenssen, Matthias Fey, Pascal Libuschewski

Neural Information Processing Systems Foundation

Motivation

Guaranteed **equivariance** and **invariance** in **capsule networks**



Equivariance of pose vectors: $L_{\mathbf{p}}(\mathbf{g} \circ \mathbf{P}, \mathbf{a}) = \mathbf{g} \circ L_{\mathbf{p}}(\mathbf{P}, \mathbf{a})$

Invariance of agreements: $L_{\mathbf{a}}(\mathbf{g} \circ \mathbf{P}, \mathbf{a}) = L_{\mathbf{a}}(\mathbf{P}, \mathbf{a})$

Equivariant routing by agreement

Given a Lie group G:

If we have a **weighted mean operator** $\mathcal{M}\colon G^n \times [0,1]^n \to G$ that is

1. left equivariant: $\mathcal{M}(\mathbf{g} \circ \mathbf{P}, \mathbf{x}) = \mathbf{g} \circ \mathcal{M}(\mathbf{P}, \mathbf{x})$
2. permutation invariant: $\mathcal{M}(\mathbf{P}, \mathbf{x}) = \mathcal{M}(\pi(\mathbf{P}), \pi(\mathbf{x}))$

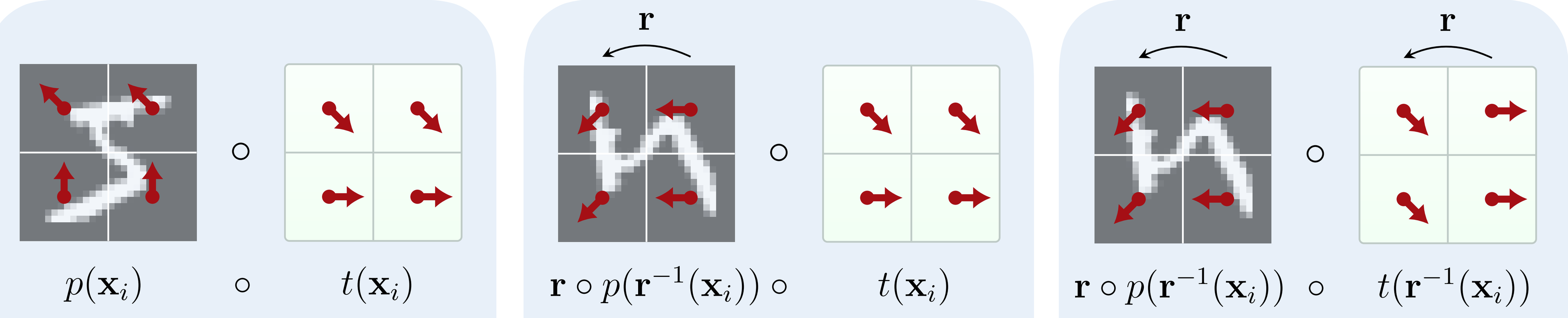
and a **distance measure** δ which is preserved by applications of $\mathbf{g} \in G$

Then: We can formulate a routing by agreement algorithm which produces equivariant poses and invariant activations

Input: poses $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_n) \in G^n$, activations $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$
Trainable parameters: transformations $\mathbf{t}_{i,j}$
Output: poses $\hat{\mathbf{P}} = (\hat{\mathbf{p}}_1, \dots, \hat{\mathbf{p}}_m) \in G^m$, activations $\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_m) \in \mathbb{R}^m$

$\mathbf{v}_{i,j} \leftarrow \mathbf{p}_i \circ \mathbf{t}_{i,j}$ — for all input capsules i and output capsules j
$\hat{\mathbf{p}}_j \leftarrow \mathcal{M}((\mathbf{v}_{1,j}, \dots, \mathbf{v}_{n,j}), \mathbf{a})$ — $\forall j$
for r iterations **do**
  $w_{i,j} \leftarrow \sigma(-\delta(\hat{\mathbf{p}}_j, \mathbf{v}_{i,j})) \cdot a_i$ — $\forall i,j$
  $\hat{\mathbf{p}}_j \leftarrow \mathcal{M}((\mathbf{v}_{1,j}, \dots, \mathbf{v}_{n,j}), \mathbf{w}_{:,j})$ — $\forall j$
end for
$\hat{a}_j \leftarrow \sigma(-\frac{1}{n}\sum_{i=1}^{n} \delta(\hat{\mathbf{p}}_j, \mathbf{v}_{i,j}))$ — $\forall j$
Return $\hat{\mathbf{p}}_1, \dots, \hat{\mathbf{p}}_m, \hat{\mathbf{a}}$

Spatial aggregation in capsule networks



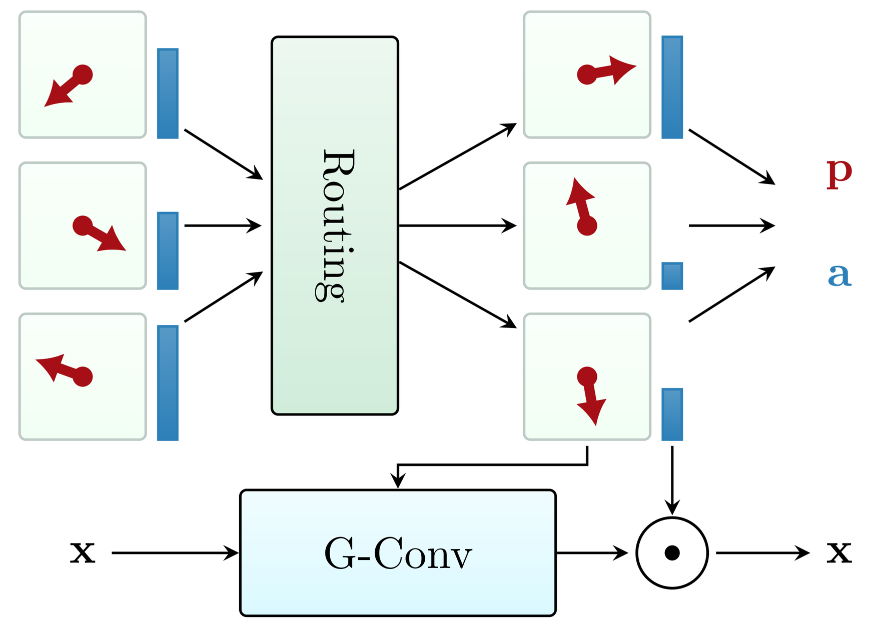
Non rotated input and poses | Rotated input, false matching | Rotated input, Pose-aligned kernels

Pose-aligning transformation kernel windows using induced representations

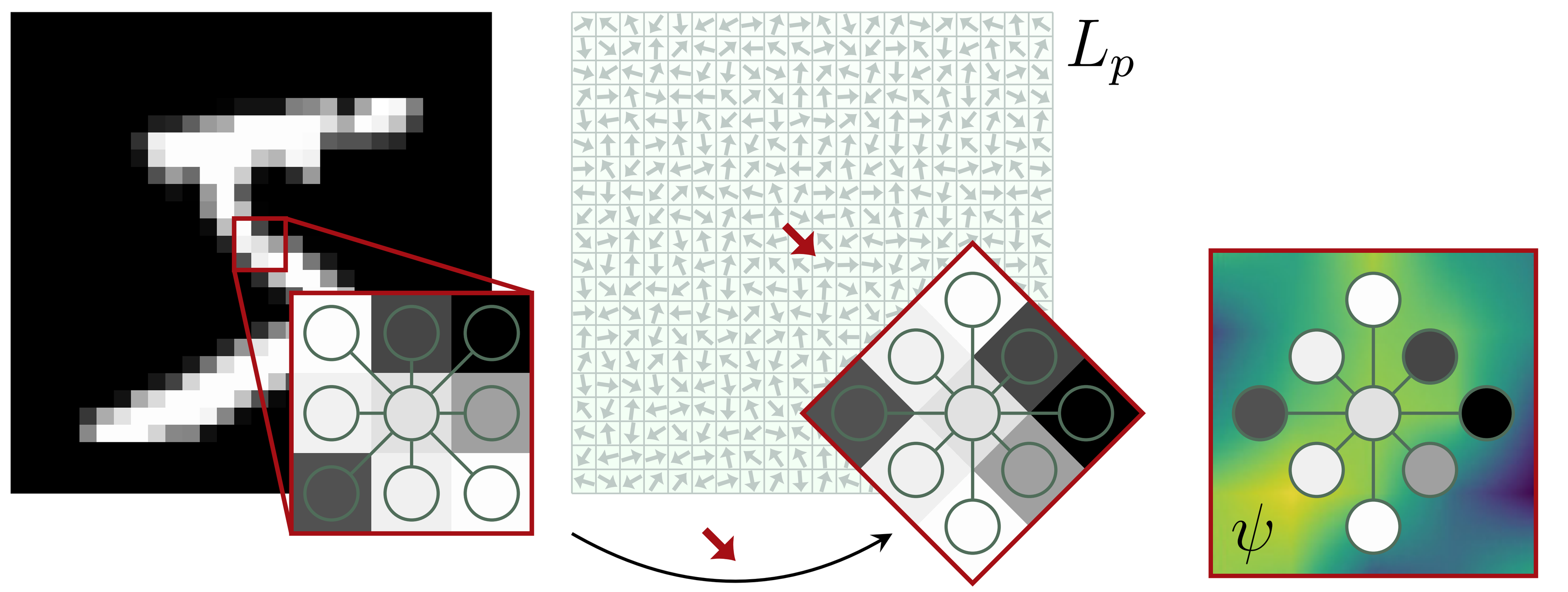
Group invariant CNNs through pose induction

Transferring invariance from capsule agreements to activations of group CNNs

$$[f \star \psi](\mathbf{g}) = \int_{\mathbf{h} \in G} f(\mathbf{h}) \psi(\mathbf{g}^{-1}\mathbf{h}) \; d\mathbf{h}$$



Realization: Using pose vector fields to transform local receptive fields



Experiments

- Proof of concept experiments on different MNIST datasets with $SO(2)$
- Perfect equi- and invariance for $\pi/2$ rotations

Results for training on rotated data

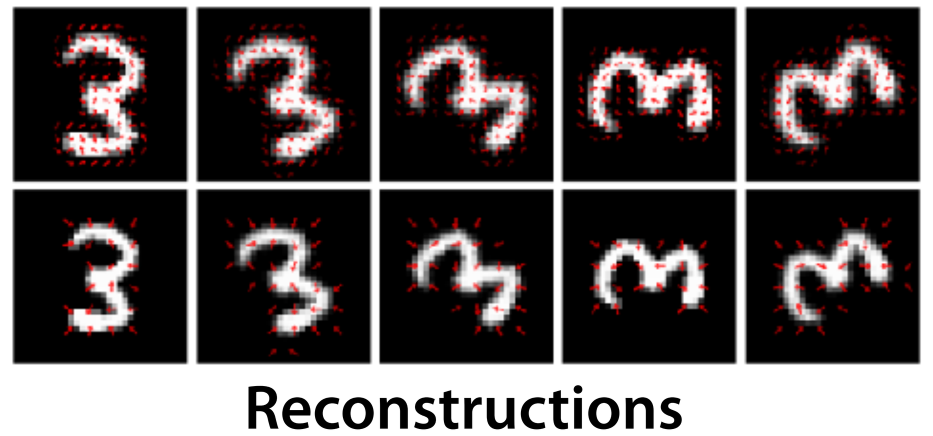
Ablation Study: Accuracy

	MNIST rot. (50k)	AffNist	MNIST rot. (10k)
CNN	92.30%	81.64%	90.19%
Capsules	94.68%	71.86%	91.87%
Whole	**98.42%**	**89.10%**	**97.40%**

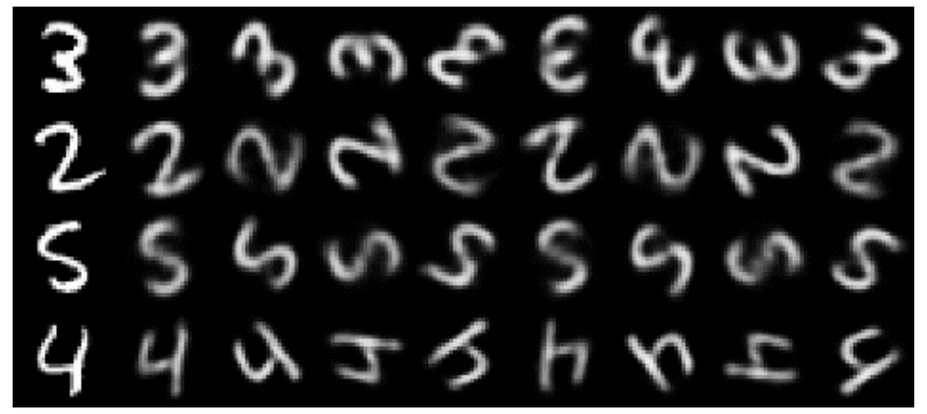
Quality of pose vectors

	Average pose error [degree]
Naive average poses	70.92
Capsules w/o recon. loss	28.32
Capsules with recon. loss	**16.21**

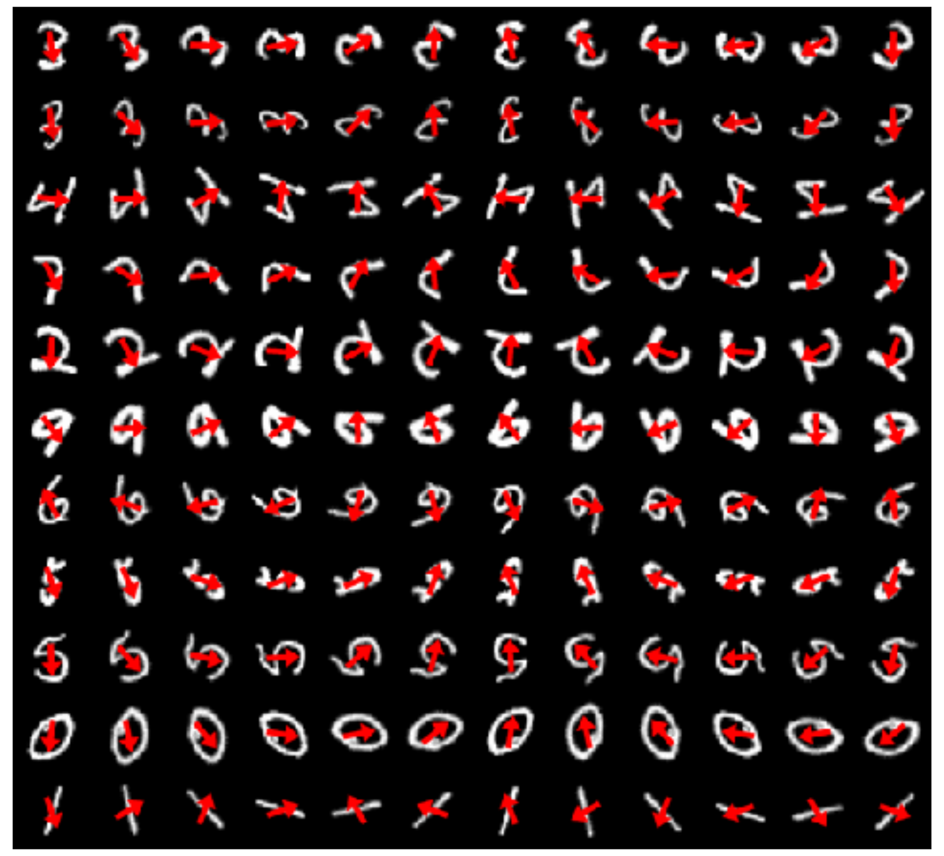
Intermediate poses



Reconstructions



Output poses



Takeaway

- We provide a mathematically grounded version of capsule networks
- Equivariance and invariance are possible under certain conditions
- Poses are useful for classification or unsupervised pose prediction tasks
- Group invariant CNNs can be sparsely evaluated by using external pose vector fields