Learning Rotation-Aware Features

From Invariant Priors to Equivariant Descriptors

TECHNISCHE UNIVERSITÄT DARMSTADT

Uwe Schmidt

Stefan Roth

Department of Computer Science
TU Darmstadt

Image Restoration

- People take pictures at portrait and landscape orientations
  - do not want this to affect image restoration

Image Restoration

- People take pictures at portrait and landscape orientations
  - do not want this to affect image restoration
- Many image priors, particularly learned ones, are not invariant to image rotation

$p\Big($  $\Big) \neq p\Big($  $\Big)$

Image Restoration

- People take pictures at portrait and landscape orientations
  - do not want this to affect image restoration
- Many image priors, particularly learned ones, are not invariant to image rotation
  - Example: effect on image denoising

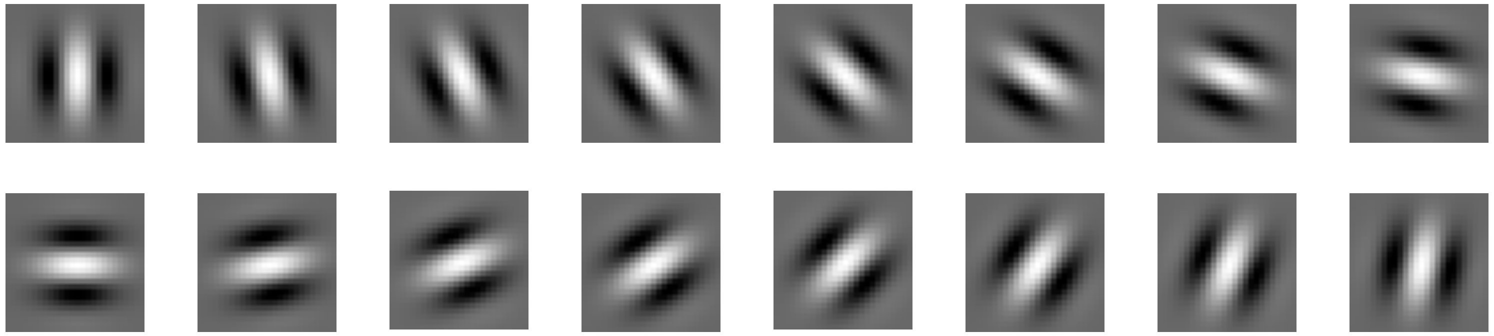
Denoised in portrait — Denoised in landscape = Noticeable difference

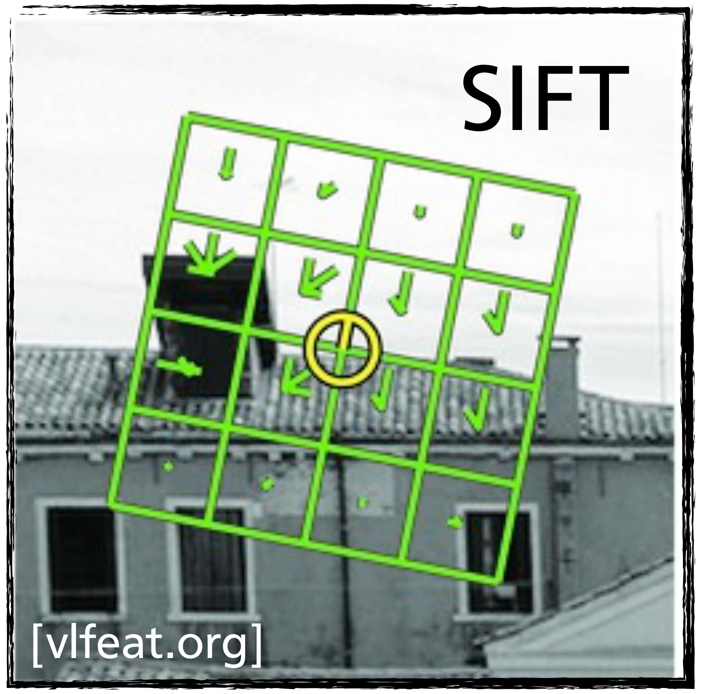
Object Detection

- Object classes may appear at arbitrary orientations in the image data (e.g. satellite images)
- Rotation-invariant image descriptor beneficial for detection
  - exist for hand-designed features (e.g. RIFF-Polar [Takacs et al., 2010])
  - not designed to work with learned image features

Local Image Features

- For both tasks, having suitable image features is important
  - e.g. filter banks
  - e.g. SIFT [Lowe, 2004], HOG [Dalal & Triggs, 2005]





- Designing good features is an important problem
  - no single best feature for all applications
  - can be non-intuitive, requires domain knowledge

Unsupervised Feature Learning

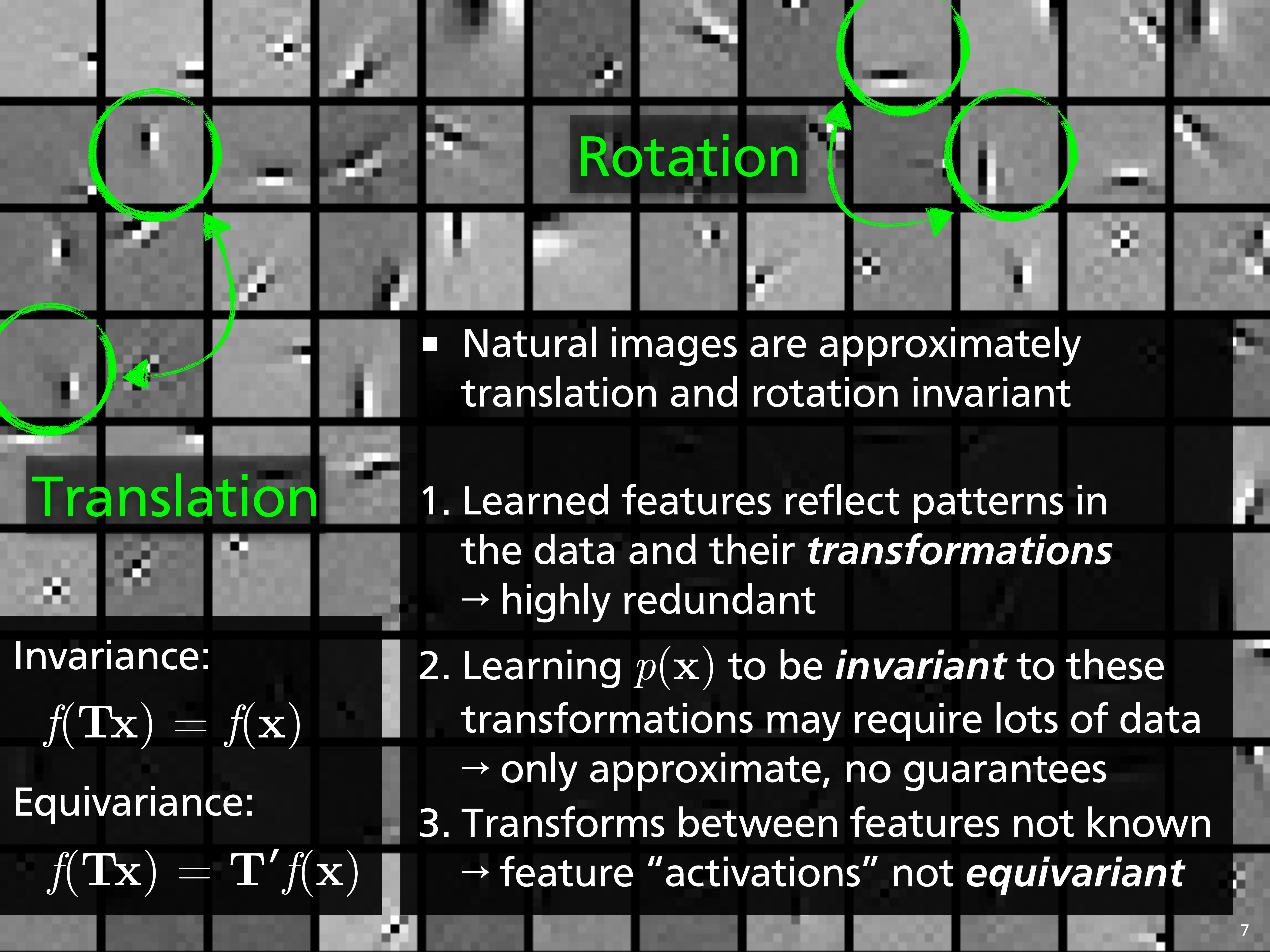
- Alternative: features learned from training data
- Tailored to specific domain
  [Hinton & Salakhutdinov, 2006; Lee et al., 2009, ...]
- Unsupervised feature learning with generative models

Features $\mathbf{F}_{(i)}$, often linear filters

$$p(\mathbf{x}) = p\left(\mathbf{x}; \mathbf{F}_{(1)}, \dots, \mathbf{F}_{(n)}\right)$$
$$= \sum_{\mathbf{h}_{(1)}, \dots, \mathbf{h}_{(n)}} p\left(\mathbf{x}, \mathbf{h}_{(1)}, \dots, \mathbf{h}_{(n)}; \mathbf{F}_{(1)}, \dots, \mathbf{F}_{(n)}\right)$$

Idea: good model of the data leads to suitable domain-adapted features

Latent variables $\mathbf{h}_{(i)}$ denote the "activations" of $\mathbf{F}_{(i)}$



- Natural images are approximately translation and rotation invariant

1. Learned features reflect patterns in the data and their ***transformations***
   → highly redundant
2. Learning $p(\mathbf{x})$ to be ***invariant*** to these transformations may require lots of data
   → only approximate, no guarantees
3. Transforms between features not known
   → feature "activations" not ***equivariant***

Invariance:

$$f(\mathbf{T}\mathbf{x}) = f(\mathbf{x})$$

Equivariance:

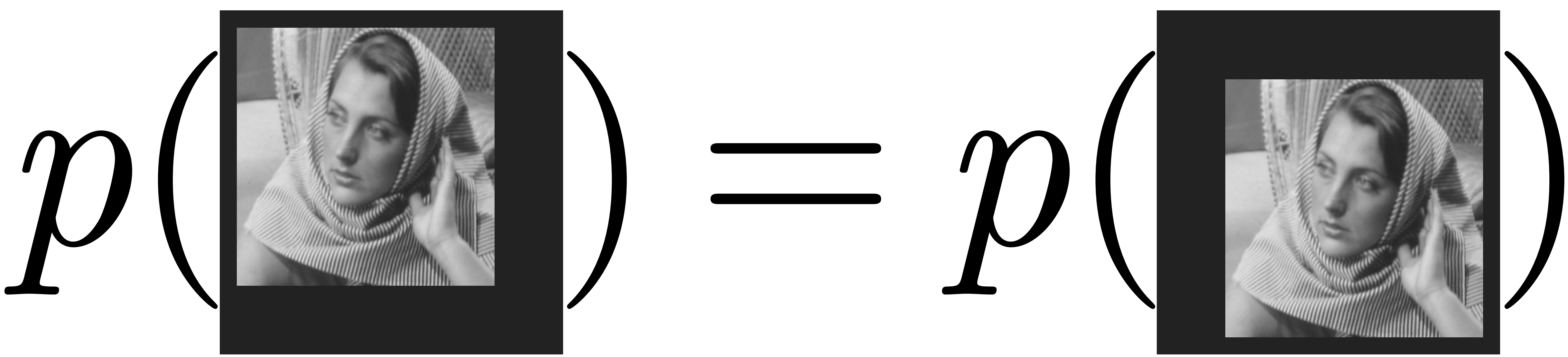
$$f(\mathbf{T}\mathbf{x}) = \mathbf{T}' f(\mathbf{x})$$

Invariances in Feature Learning

Translation invariance of $p(\mathbf{x})$ often addressed through convolutional model [Norouzi et al., 2009; standard MRFs]

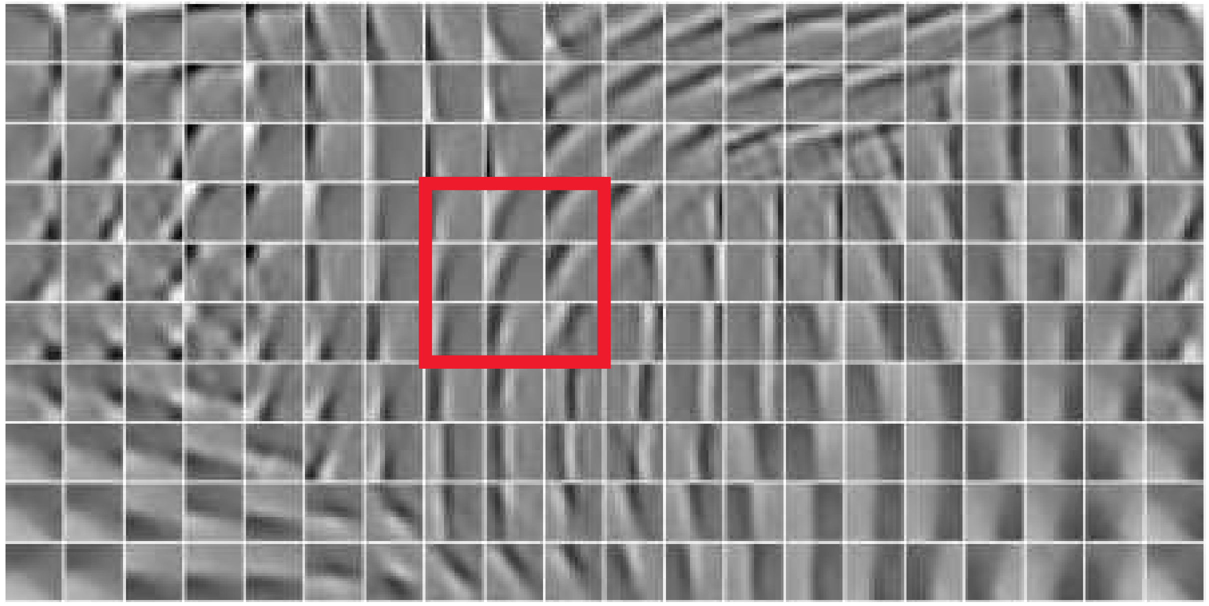
- features applied convolutionally at all locations of the image
- requires many fewer features:
  redundancies due to translations are suppressed

- Typical MRF-based image priors are translation invariant



Transformations between Features

- Topographic filter maps group similar features [Kavukcuoglu et al., 2009]
  - However: transformations between features unknown
  - Learned features reflect patterns and their transformations



[Kavukcuoglu et al., 2009]

- Idea: Explicitly build transformations into the model to suppress redundancies in learned features
  - Transformed features implicitly added with known relationship
  - Convolutional models do this for translations
- Here: Transformations beyond translation, focus on rotation

Product Models

- Linear image features
- (Non-linear) factors

$$p(\mathbf{x}) \propto \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\mathbf{F}_{(i)}\mathbf{x}; \theta_i\right)$$

PoE	MRF	PCA	RBM	ICA
Product of Experts [Hinton, 2002]	Markov Random Field [Li, 2001]	Principal Component Analysis	Restricted Boltzmann Machine [Hinton, 2002]	Independent Component Analysis

Invariant Product Models

TECHNISCHE UNIVERSITÄT DARMSTADT

- Approach: explicitly build transformations into the model
  - separate them from the pattern in the data
  - model guaranteed to be invariant to built-in transformations
  - frugal model representation due to parameter sharing

$$p_{\mathcal{T}}(\mathbf{x}) \propto \prod_{i=1}^{|\mathcal{F}|} \prod_{j=1}^{|\mathcal{T}|} \phi_i\left(\mathbf{F}_{(i)} \mathbf{T}_{(j)} \mathbf{x}; \theta_i\right)$$

Feature $\mathbf{F}_{(i)}$ is used with all transformations $\mathbf{T}_{(j)}$

$$p(\mathbf{x}) \propto \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\mathbf{F}_{(i)} \mathbf{x}; \theta_i\right)$$

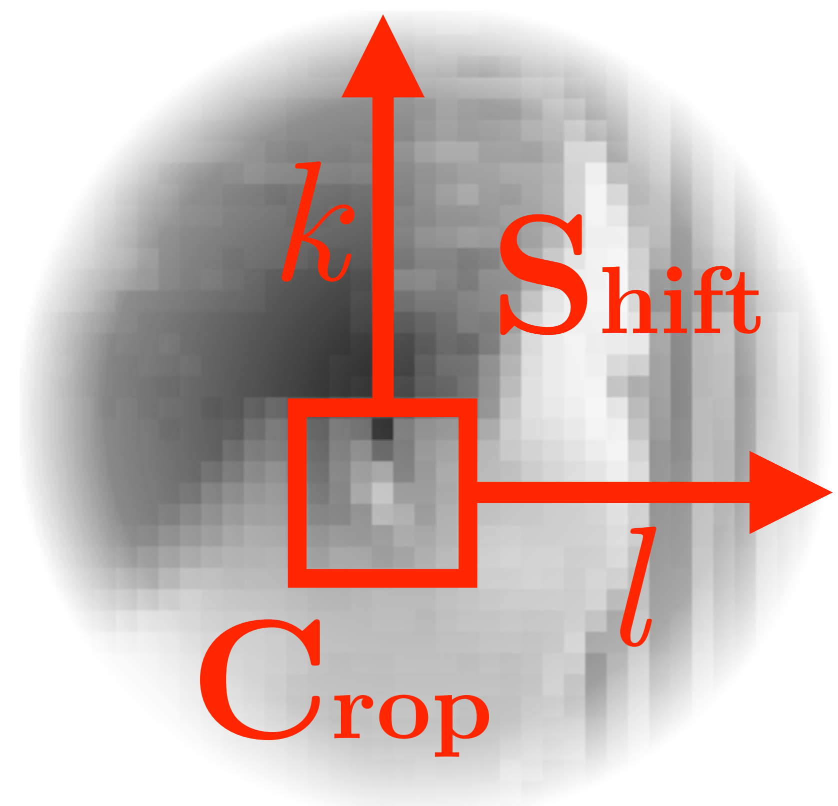
Linear Transformation Sets

$$p_{\mathcal{T}}(\mathbf{x}) \propto \prod_{i=1}^{|\mathcal{F}|} \prod_{j=1}^{|\mathcal{T}|} \phi_i\left(\mathbf{F}_{(i)} \mathbf{T}_{(j)} \mathbf{x}; \theta_i\right)$$

"Convolutional" transformations → MRFs

$$\mathcal{T}_{\mathrm{C}} = \left\{\mathbf{C} \cdot \mathbf{S}_{(k,l)}\right\}$$

$$\mathcal{T}_{\mathrm{RC}} = \left\{\mathbf{R}_{(\omega)} \cdot \mathbf{C} \cdot \mathbf{S}_{(k,l)}\right\}$$



Rotate

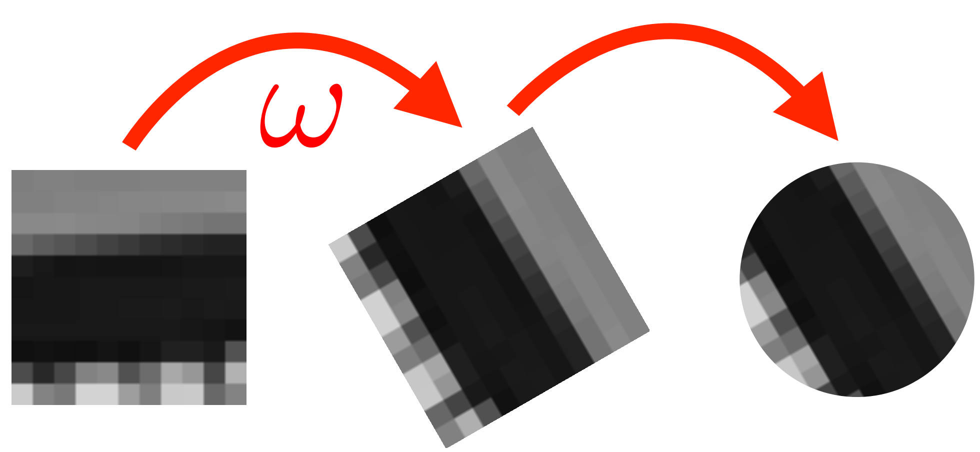


image rotation (bilinear interpolation)

Benefits of Invariant Product Model

- Existing learning and inference needs little modification
  - transformed features implicitly defined with shared parameters
- All transformations treated equally
  - e.g. in contrast to [Kivinen & Williams, 2011]

- “Transformation-aware” learning
  - feature redundancies suppressed
- Not limited to specific models or transformations

- To come: equivariant image descriptor with learned features

Learning Rotation-invariant Image Priors

TECHNISCHE UNIVERSITÄT DARMSTADT

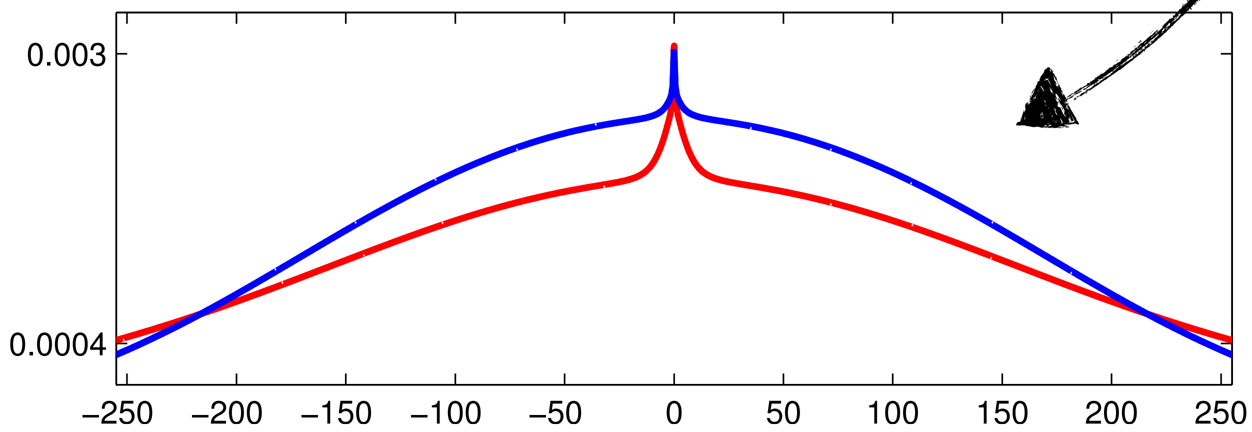
- Motivation: rotation-equivariant image restoration
- Extend Fields of Experts (FoE) image prior [Roth & Black, 2009, Schmidt et al., 2010] to be invariant to multiples of 90° rotations

$$p_{\text{R-FoE}}(\mathbf{x}) \propto \prod_{\omega \in \Omega} \prod_{(k,l)} \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\mathbf{J}_i^{\mathrm{T}} \cdot \mathbf{R}_{(\omega)} \cdot \mathbf{C} \cdot \mathbf{S}_{(k,l)} \mathbf{x}; \theta_i\right)$$

Learning Rotation-invariant Image Priors

- Motivation: rotation-equivariant image restoration
- Extend Fields of Experts (FoE) image prior [Roth & Black, 2009, Schmidt et al., 2010] to be invariant to multiples of 90° rotations

$$p_{\text{R-FoE}}(\mathbf{x}) \propto \prod_{\omega \in \Omega} \prod_{(k,l)} \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\mathbf{J}_i^{\mathrm{T}} \cdot \mathbf{R}_{(\omega)} \cdot \mathbf{C} \cdot \mathbf{S}_{(k,l)} \mathbf{x}; \theta_i\right)$$

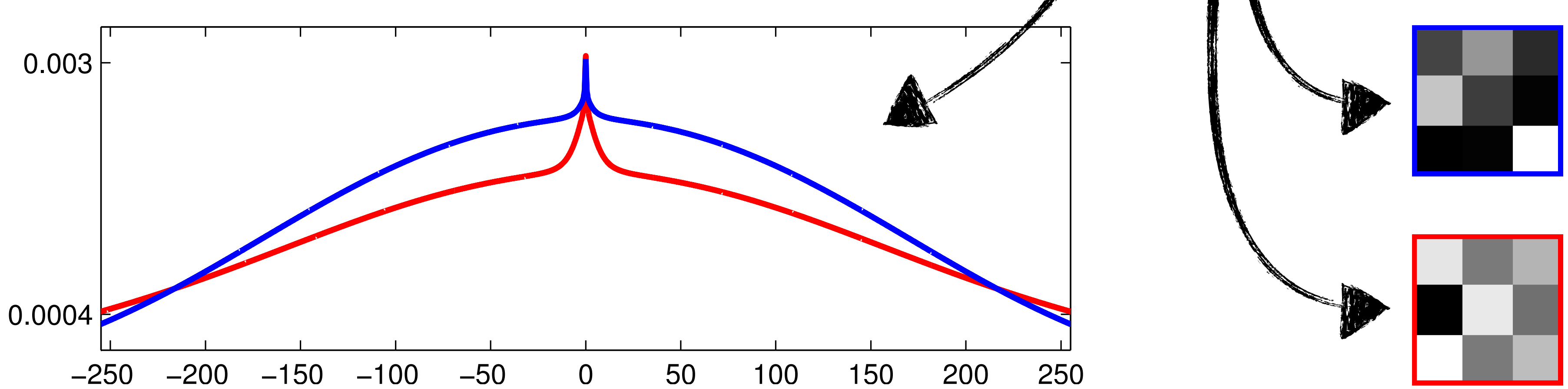


Learning Rotation-invariant Image Priors

TECHNISCHE UNIVERSITÄT DARMSTADT

- Motivation: rotation-equivariant image restoration
- Extend Fields of Experts (FoE) image prior [Roth & Black, 2009, Schmidt et al., 2010] to be invariant to multiples of 90° rotations

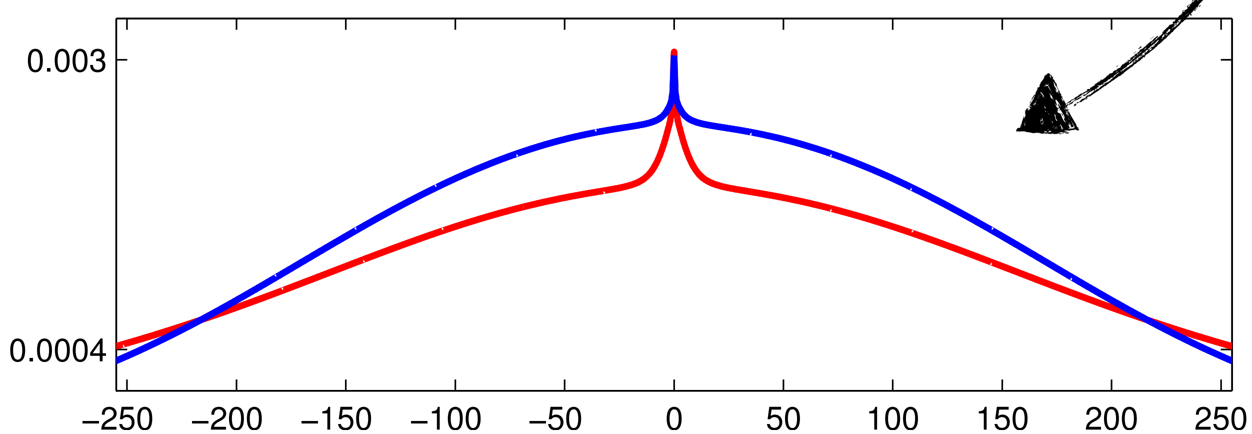
$$p_{\text{R-FoE}}(\mathbf{x}) \propto \prod_{\omega \in \Omega} \prod_{(k,l)} \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\mathbf{J}_i^{\mathrm{T}} \cdot \mathbf{R}_{(\omega)} \cdot \mathbf{C} \cdot \mathbf{S}_{(k,l)} \mathbf{x}; \theta_i\right)$$

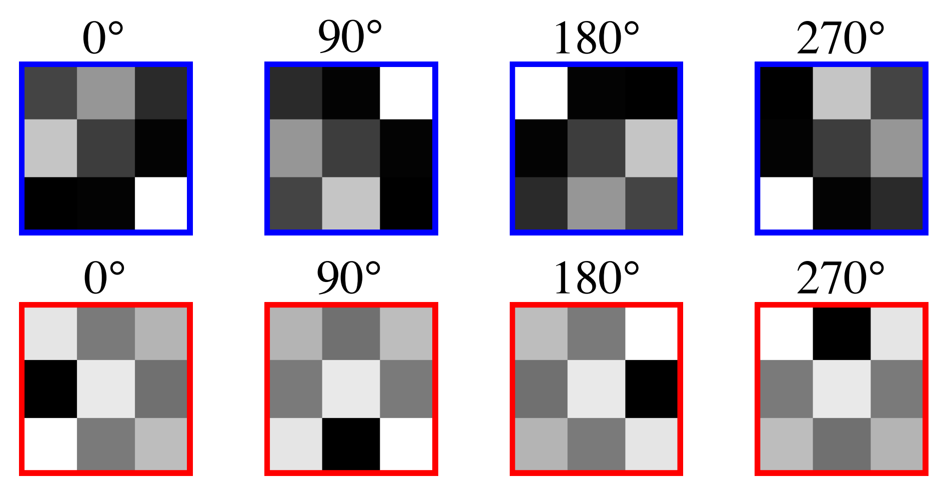


Learning Rotation-invariant Image Priors

- Motivation: rotation-equivariant image restoration
- Extend Fields of Experts (FoE) image prior [Roth & Black, 2009, Schmidt et al., 2010] to be invariant to multiples of 90° rotations

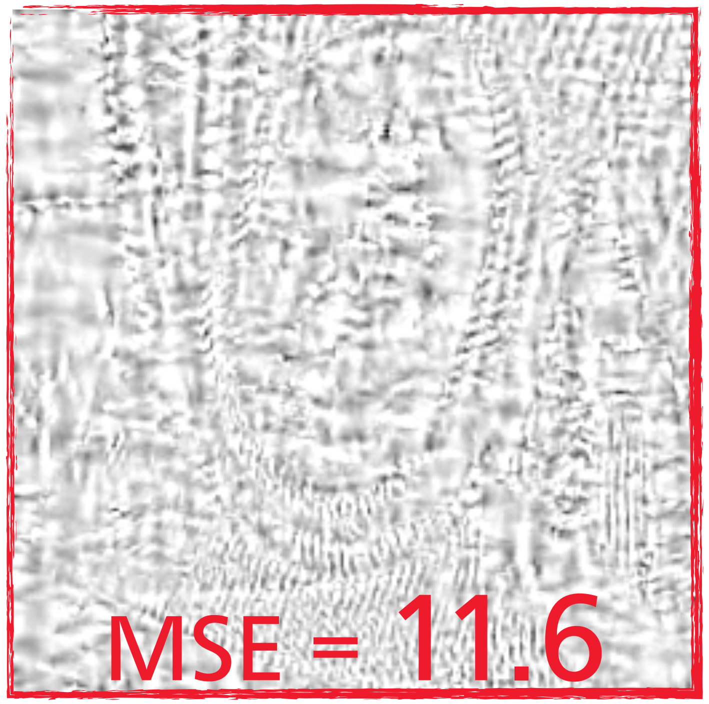
$$p_{\text{R-FoE}}(\mathbf{x}) \propto \prod_{\omega\in\Omega}\prod_{(k,l)}\prod_{i=1}^{|\mathcal{F}|}\phi_i\left(\mathbf{J}_i^{\mathrm{T}}\cdot\mathbf{R}_{(\omega)}\cdot\mathbf{C}\cdot\mathbf{S}_{(k,l)}\mathbf{x};\theta_i\right)$$



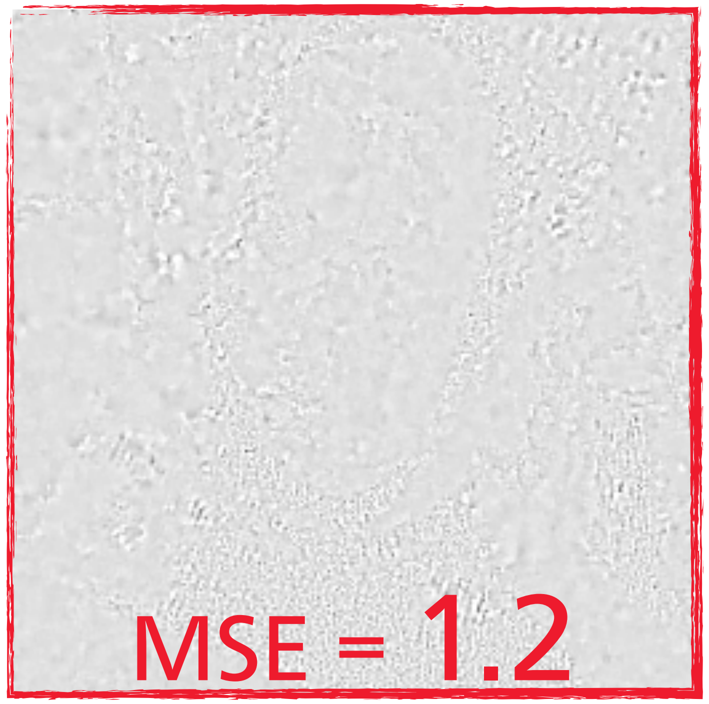


Rotation-equivariant Denoising

Differences between denoising image in portrait or landscape orientation

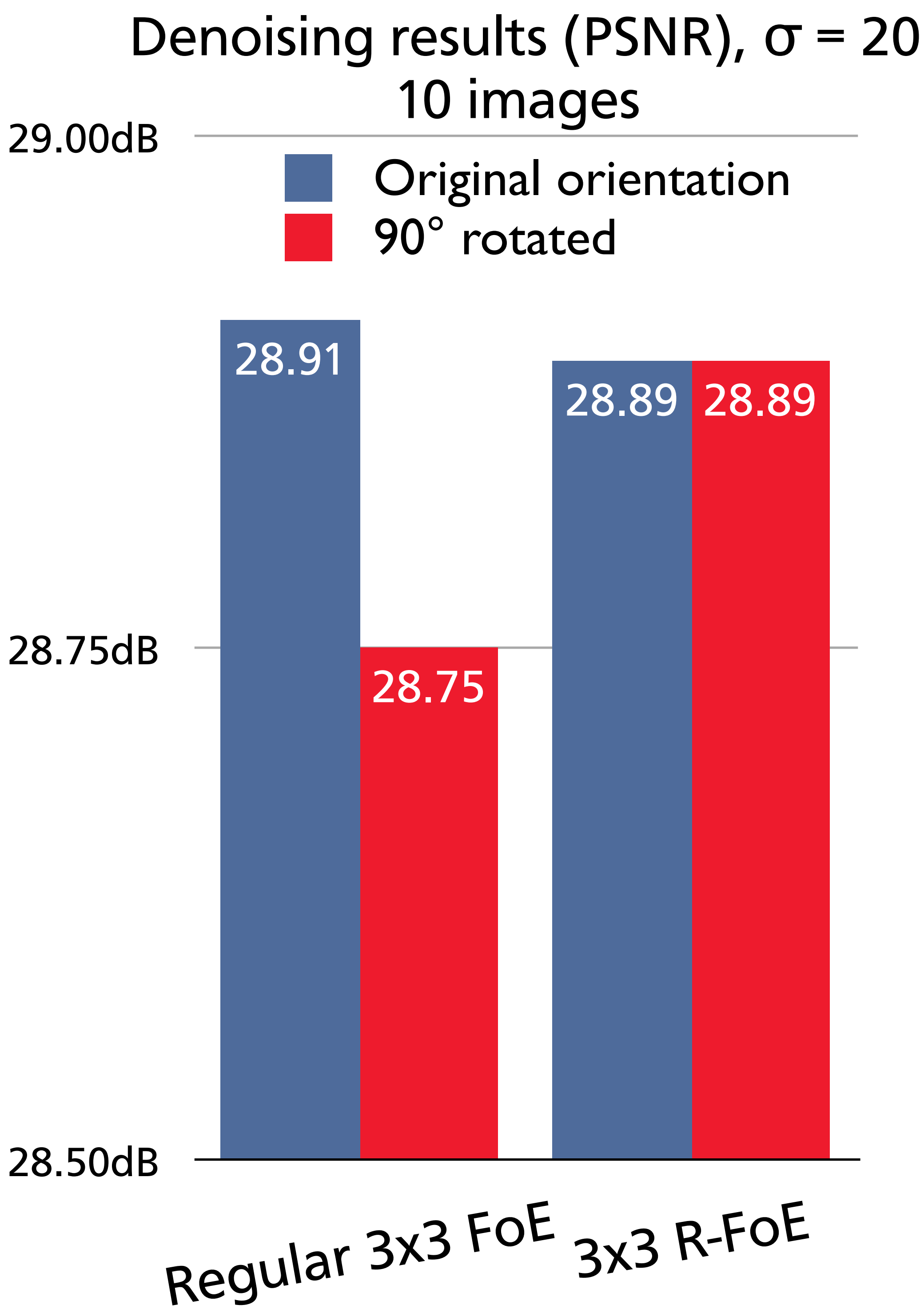


Regular 3x3 FoE [Schmidt et al., 2010]



Rotation-invariant 3x3 R-FoE

Proposed R-FoE allows rotation-equivariant image restoration without sacrificing performance



Learning Rotation-aware Image Features

- Motivation: rotation-equiv. features for object detection
- Extend convolutional RBM [Lee et al., 2009; Norouzi et al., 2009] to be invariant to multiples of 45° rotations (RC-RBM)

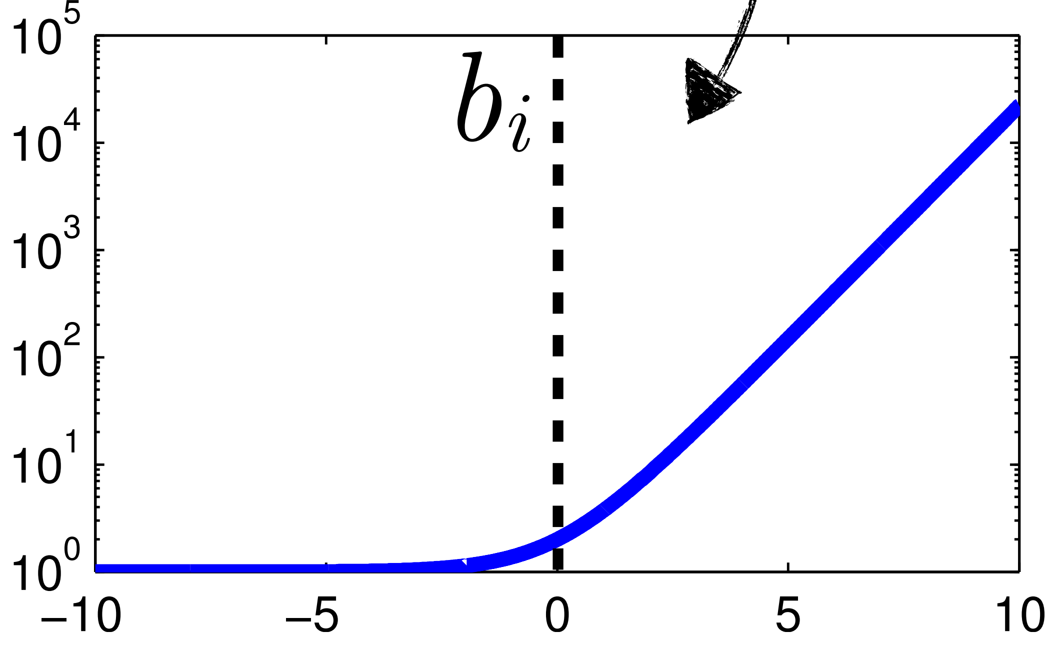
$$p_{\text{RC-RBM}}(\mathbf{x}) \propto \exp\left(\mathbf{c}^{\mathrm{T}}\mathbf{x}\right) \cdot \prod_{\omega \in \Omega} \prod_{(k,l)} \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\mathbf{w}_i^{\mathrm{T}} \cdot \mathbf{R}_{(\omega)} \cdot \mathbf{C} \cdot \mathbf{S}_{(k,l)}\mathbf{x}; b_i\right)$$

Learning Rotation-aware Image Features

TECHNISCHE UNIVERSITÄT DARMSTADT

- Motivation: rotation-equiv. features for object detection
- Extend convolutional RBM [Lee et al., 2009; Norouzi et al., 2009] to be invariant to multiples of 45° rotations (RC-RBM)

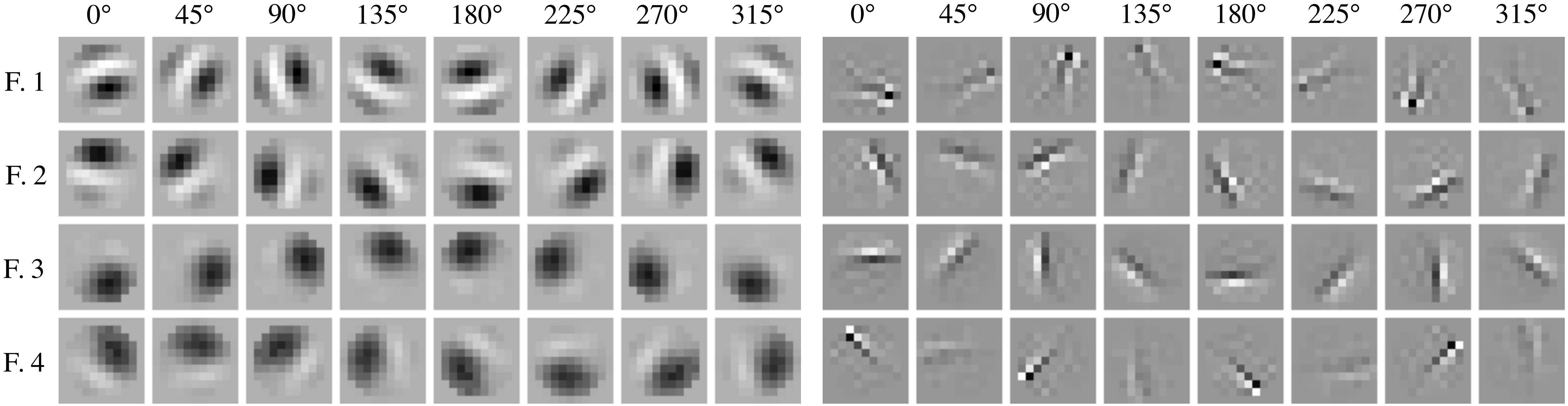
$$p_{\text{RC-RBM}}(\mathbf{x}) \propto \exp\left(\mathbf{c}^{\mathrm{T}}\mathbf{x}\right) \cdot \prod_{\omega \in \Omega} \prod_{(k,l)} \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\mathbf{w}_i^{\mathrm{T}} \cdot \mathbf{R}_{(\omega)} \cdot \mathbf{C} \cdot \mathbf{S}_{(k,l)}\mathbf{x}; b_i\right)$$



Learning Rotation-aware Image Features

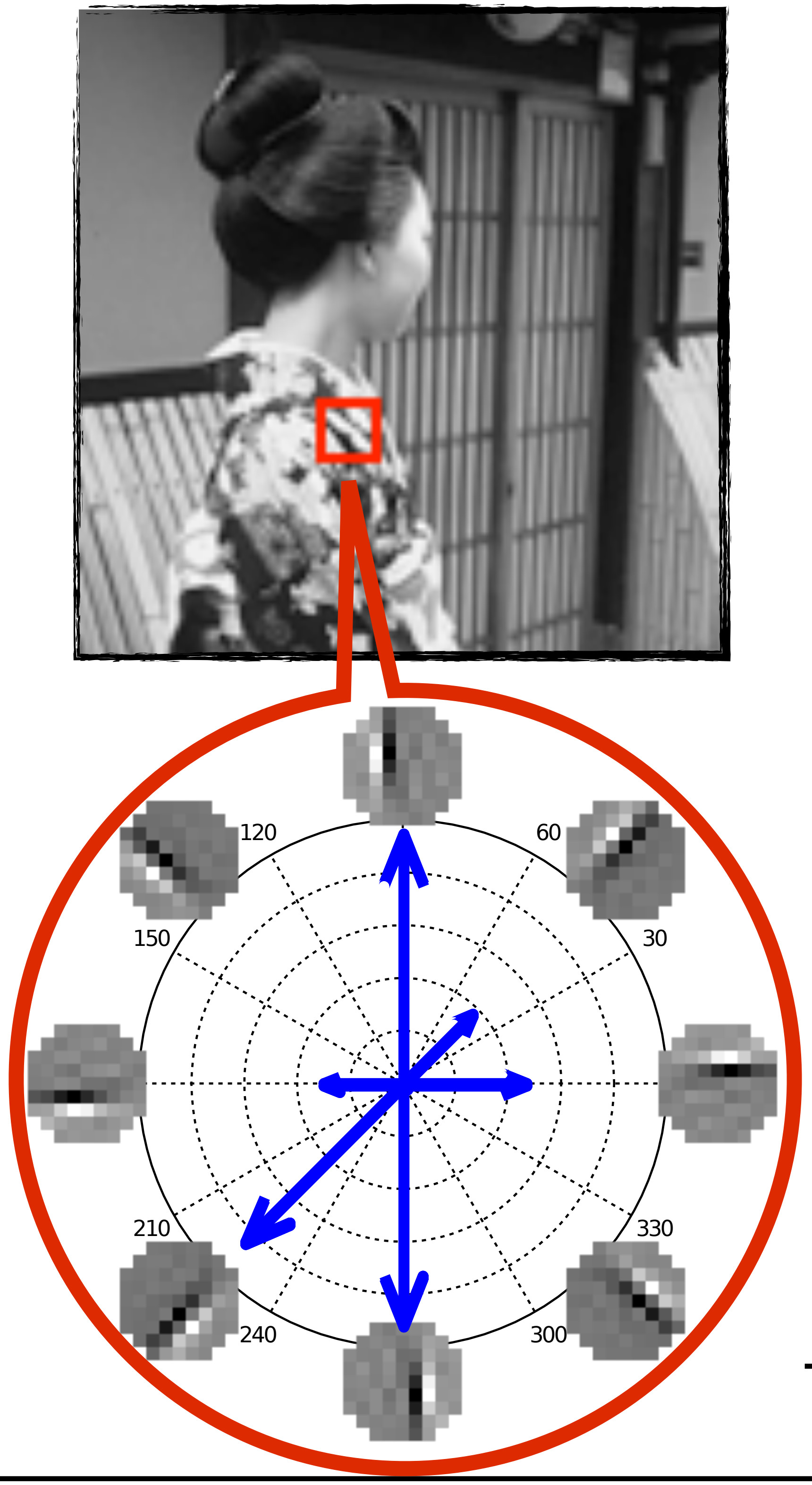
- Motivation: rotation-equiv. features for object detection
- Extend convolutional RBM [Lee et al., 2009; Norouzi et al., 2009] to be invariant to multiples of 45° rotations (RC-RBM)

$$p_{\text{RC-RBM}}(\mathbf{x}) \propto \exp\left(\mathbf{c}^{\mathrm{T}}\mathbf{x}\right) \cdot \prod_{\omega\in\Omega} \prod_{(k,l)} \prod_{i=1}^{|\mathcal{F}|} \phi_i\left(\boxed{\mathbf{w}_i^{\mathrm{T}} \cdot \mathbf{R}_{(\omega)}} \cdot \mathbf{C} \cdot \mathbf{S}_{(k,l)}\mathbf{x}; b_i\right)$$



MNIST handwritten digits Natural images (whitened)

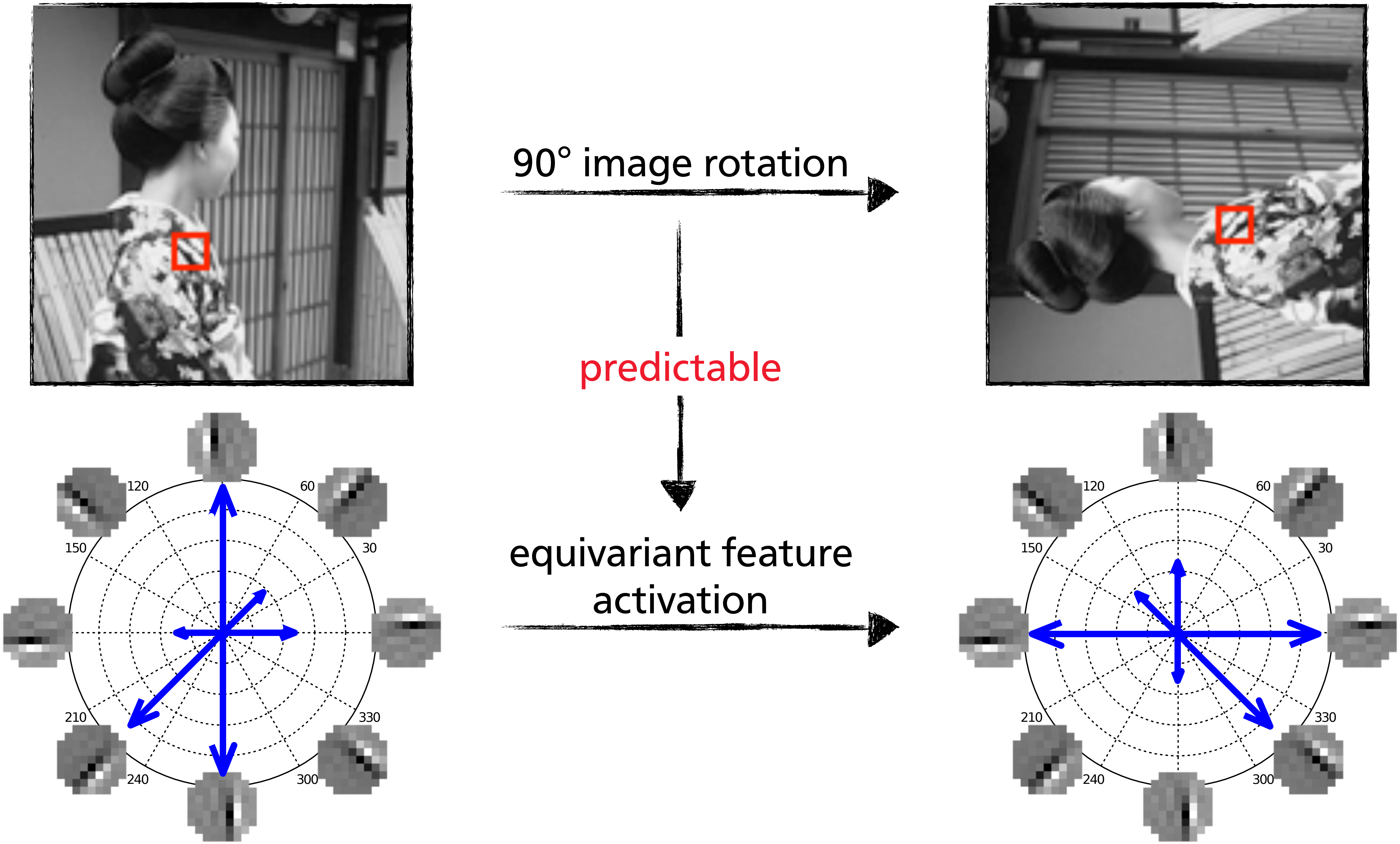
Locally Equivariant Feature Activation



Feature activation
for image patch in red

Locally Equivariant Feature Activation

TECHNISCHE
UNIVERSITÄT
DARMSTADT

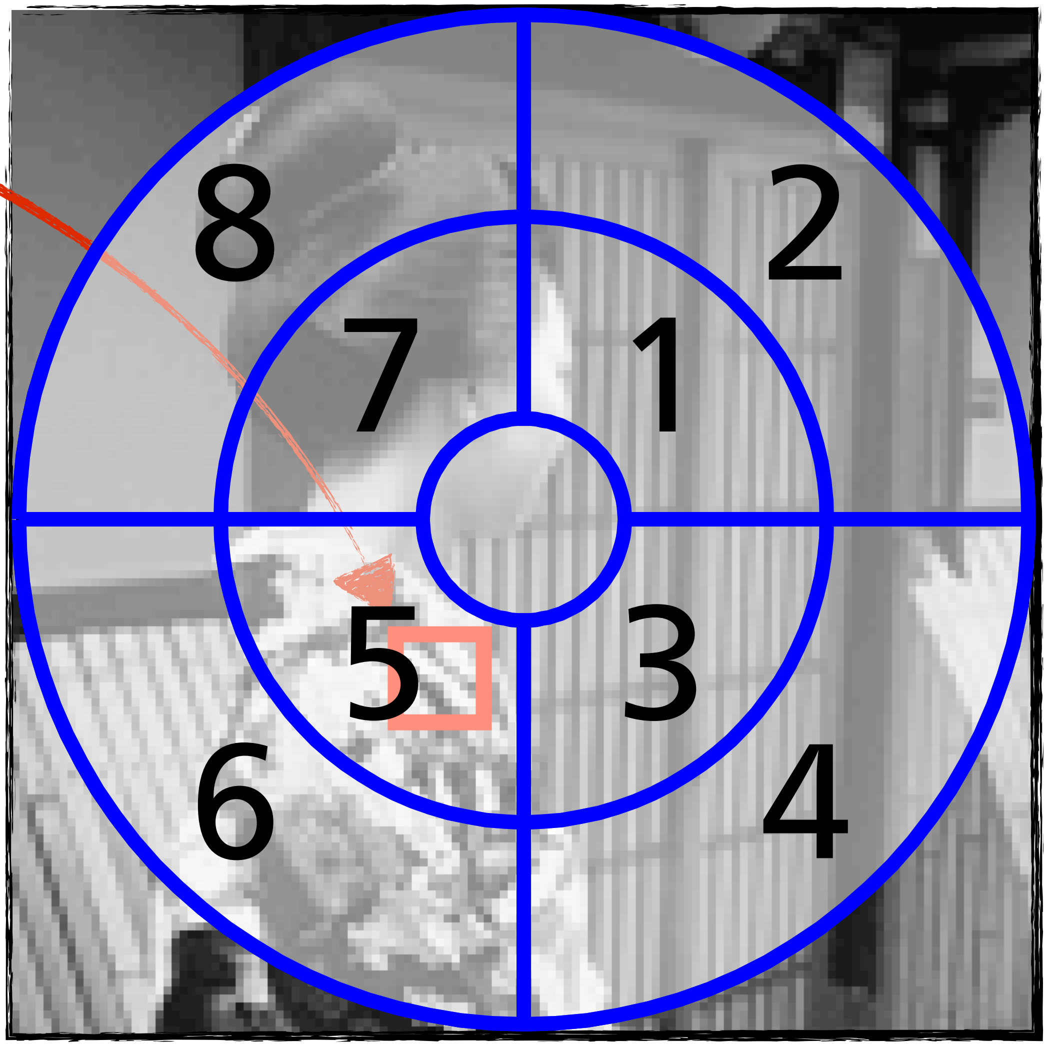


From Local to Global Equivariance

TECHNISCHE UNIVERSITÄT DARMSTADT

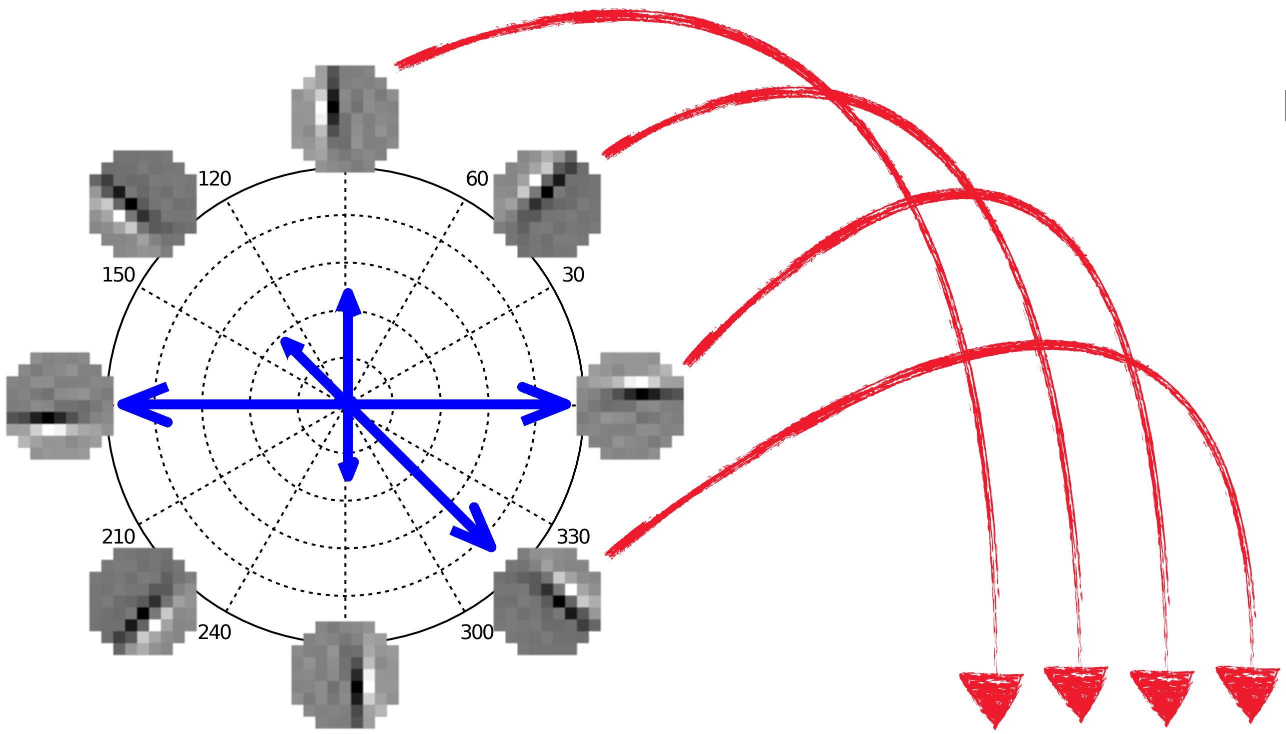
- *Here*: globally rotation-equivariant image descriptor based on (learned) locally rotation-equivariant features

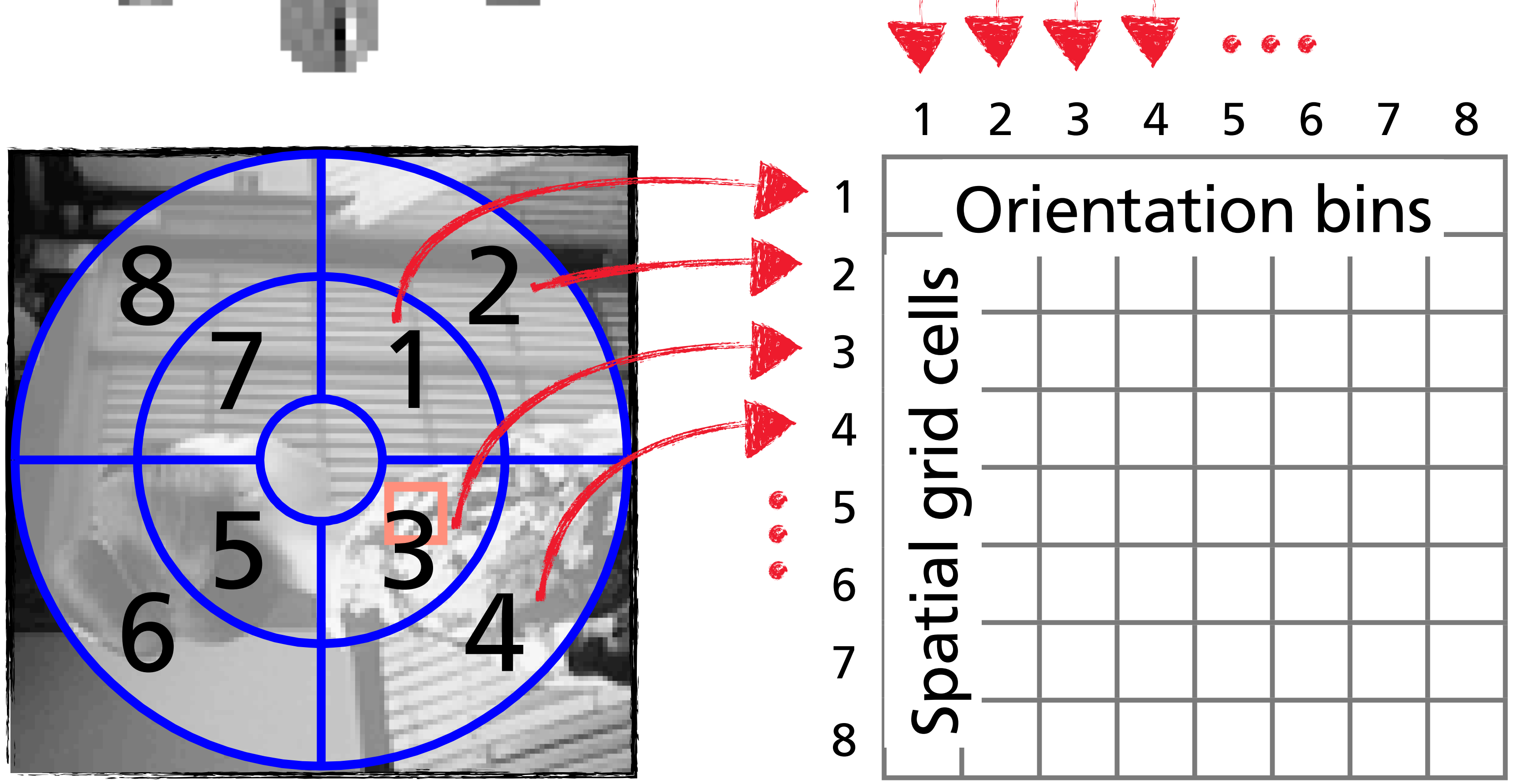
1. Dense feature extraction (at all orientations)
2. Build spatial histograms on a polar grid
3. Arrange histograms in a matrix: image rotations ↔ cyclical shifts
   - rotation-*equivariant* descriptor
4. *Optional*: Use magnitude of 2D Fourier transform to gain a rotation-*invariant* descriptor



Descriptor Design

TECHNISCHE
UNIVERSITÄT
DARMSTADT





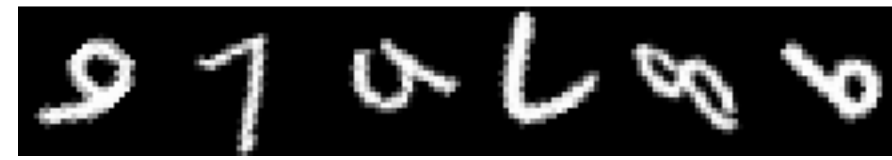
- **E**quivariant **H**istogram of **O**riented **F**eatures (**EHOF**)
  - Image rotation → cyclical translation of 2D descriptor matrix
- **I**nvariant **H**istogram of **O**riented **F**eatures (**IHOF**)
  - Magnitude of matrix's 2D Fourier transform invariant to cyclical translations

MNIST Digit Recognition

- Widely used benchmark for feature learning

MNIST

MNIST-rot



- *MNIST*: RC-RBM EHOF competitive with multilayer deep networks
- *MNIST-rot*: RC-RBM IHOF outperforms existing approaches

	Test error	
Model / Features	MNIST	MNIST-rot
RC-RBM EHOF	0.85%	6.36%
RC-RBM IHOF	2.66%	5.47%
Multilayer C-RBM [Lee et al.]	0.82%	–
SDAIC [Larochelle et al.]	–	8.07%

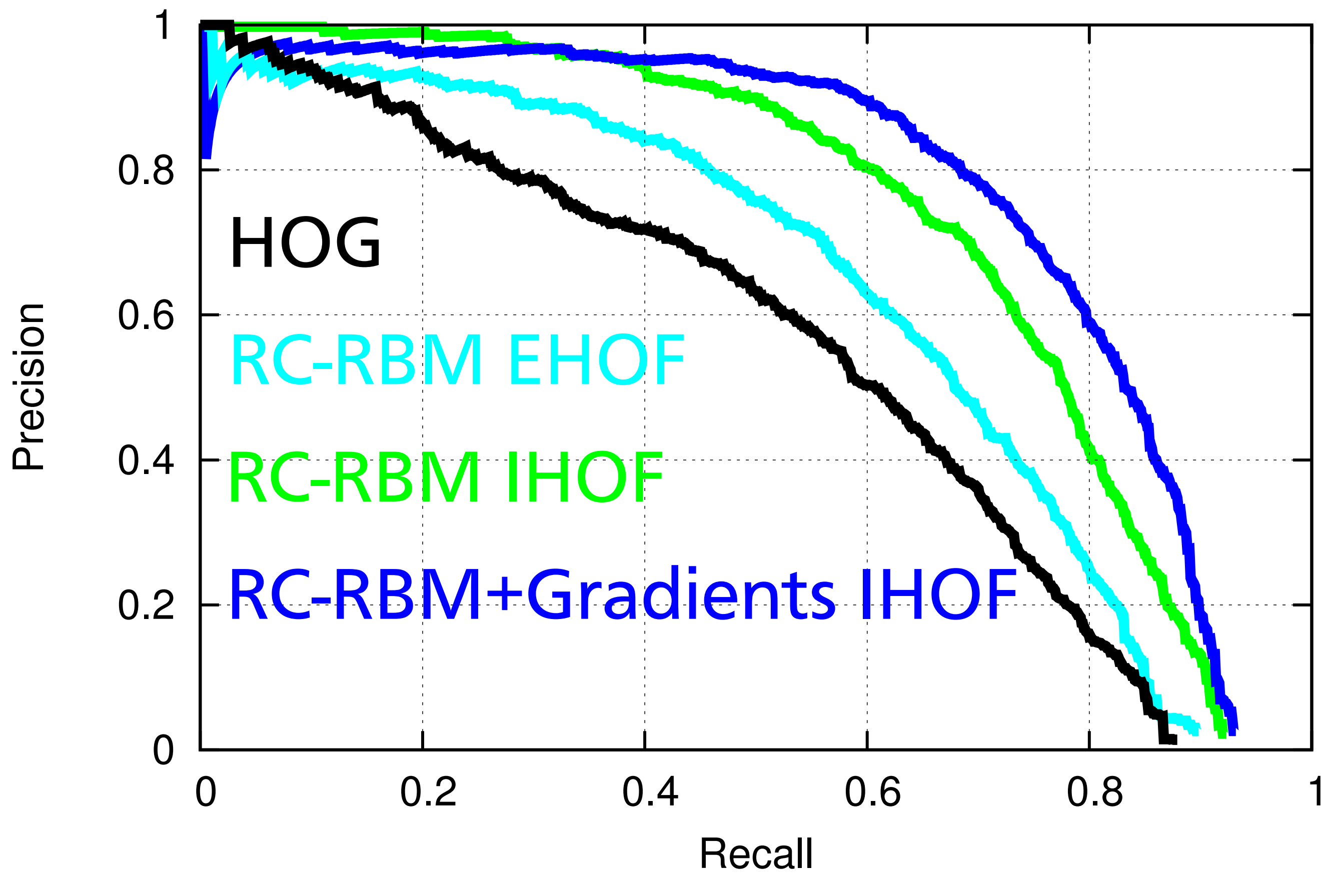
Rotation-invariant Car Detection

TECHNISCHE UNIVERSITÄT DARMSTADT

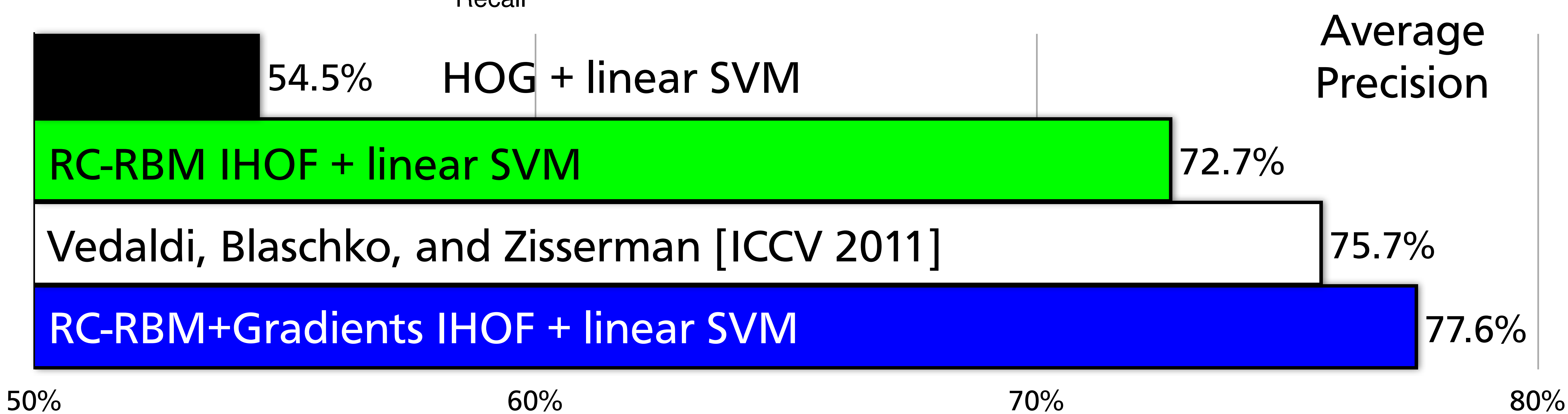
- Car detection in satellite images [Heitz & Koller, 2008]
- Cars occur at arbitrary orientations
- Sliding-window detector (linear SVM)
- EHOF/IHOF descriptor with RC-RBM features
  - optionally: image gradients

Car Detection Results

TECHNISCHE UNIVERSITÄT DARMSTADT



Our approach:
Simple classifier +
invariant features



Summary

Code for EHOF/IHOF descriptor available soon

- Learned rotation-aware features
  - based on framework for transformation-invariant product models
  - generalizes existing convolutional models to broader classes of linear transformations
  - frugal w.r.t. parameters and training data
- Learned rotation-invariant image prior
  - demonstrated rotation-equivariant image restoration
- Learned locally rotation-equivariant features (based on RBM)
  - designed globally rotation-equivariant image descriptor
  - showed rotation-invariant object recognition and detection