TRULY SCALE-EQUIVARIANT DEEP NETS WITH FOURIER LAYERS

Md Ashiqur Rahman Raymond A. Yeh

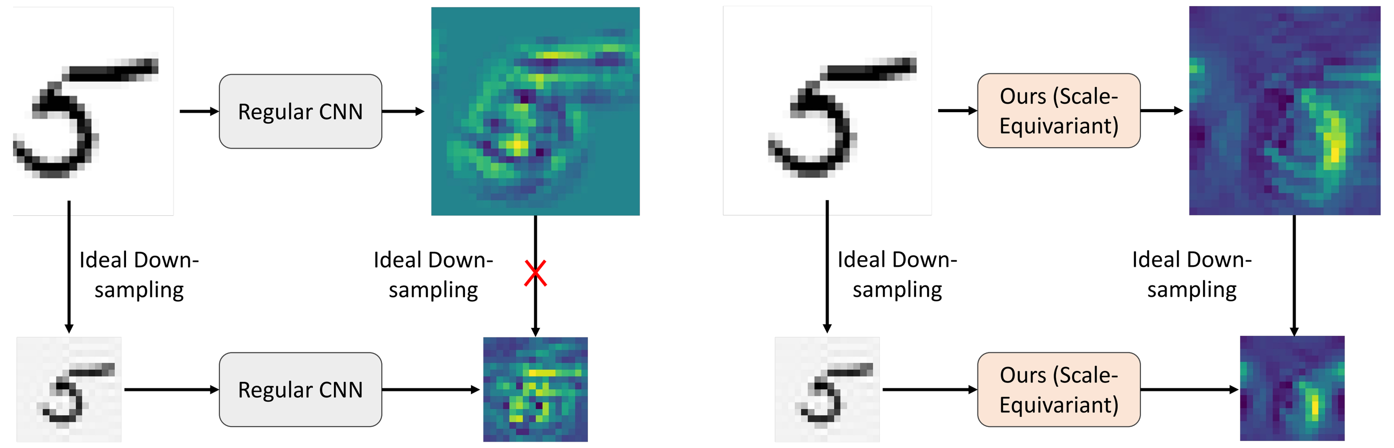
Department of Computer Science, Purdue University

PURDUE UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

MOTIVATION

Scale Equivariance:



- Scale-equivariance is crucial for consistent performance
- Prior works did not consider aliasing, resulting in equivariance error

Can we design a perfect scale equivariant layer?

- Maintain scale consistency while achieving good performance

Contributions:

- Propose scale-equivariant Fourier Layer
- End-to-end equivariance with non-linearity and pooling
- Connect scale equivariance to classification task for improved scale consistency.

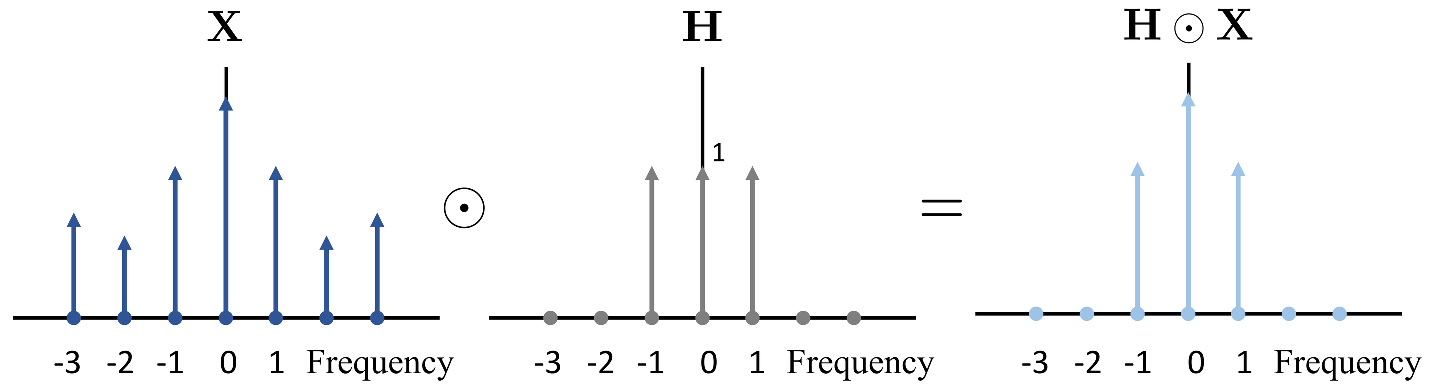
INTRODUCTION

Scaling Operation: Down-scaling a signal $\mathbf{x} \in \mathbb{R}^N$ can be performed by subsampling by a scaling factor, i.e.,

$$\texttt{Sub}_R(\mathbf{x})[n] = \mathbf{x}[Rn].$$

Ideal Down-sampling: To avoid aliasing from subsampling, a low-pass filter h must be performed, *i.e.*,

$$\mathcal{D}_R(\mathbf{x}) = \texttt{Sub}_R(\mathbf{h} \circledast \mathbf{x}).$$



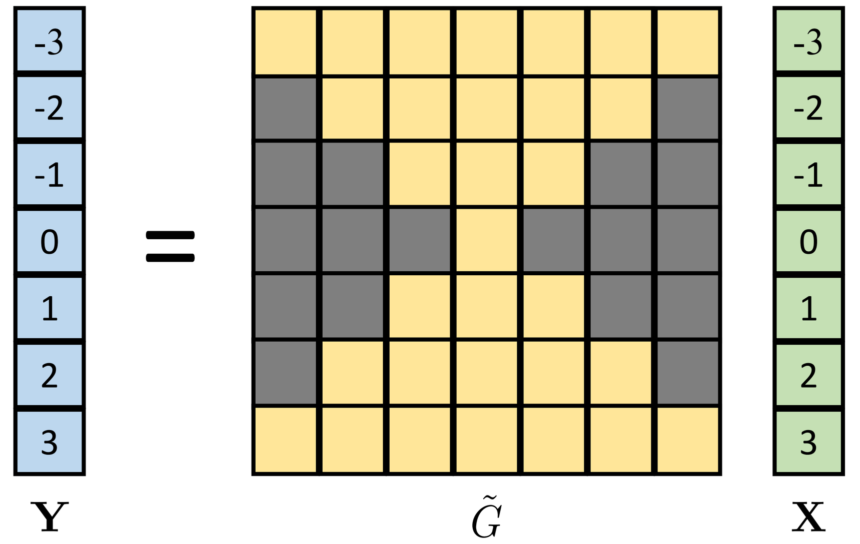
OUR APPROACH

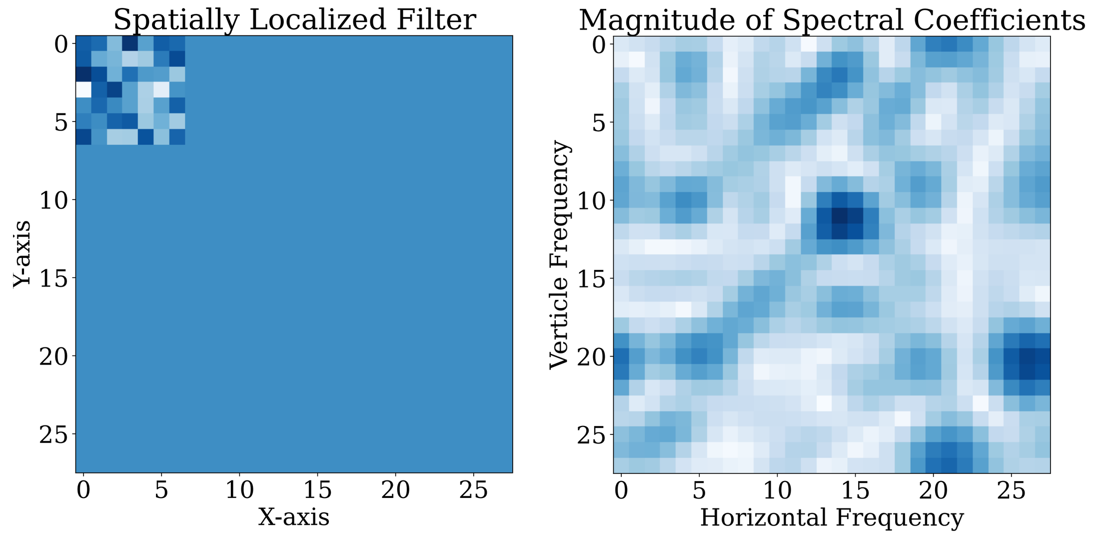
Scale Equivariant Deep Nets

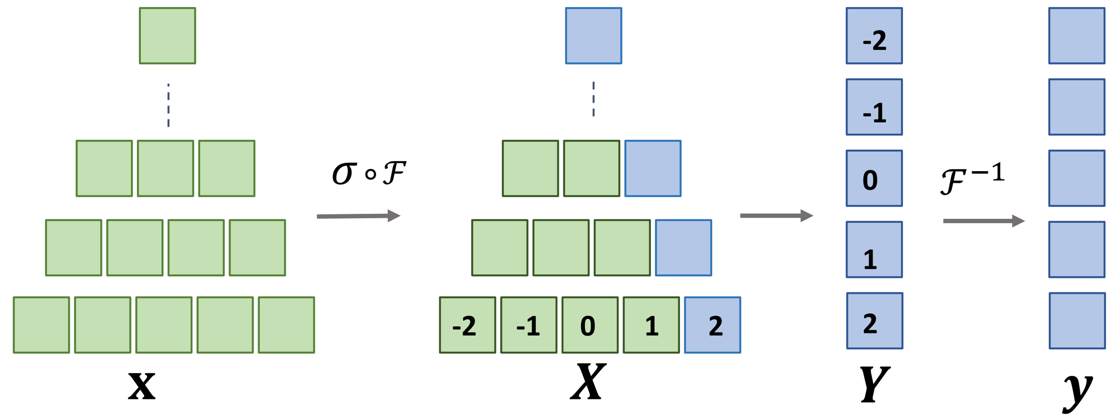
Let g denote a deep net such that $\mathbf{y} = g(\mathbf{x})$. If this deep net g can be equivalently represented as a set of functions $\tilde{G}_k : \mathbb{C}^{2k+1} \to \mathbb{C}$ such that

$$\mathbf{Y}[k] = \tilde{G}_k(\mathbf{X}[-k:k]) \;\; \forall k$$

then g is scale-equivariant. In other words, the output's frequency terms can only depend on the terms in $\mathbf{X}$ that are **equal or lower** in frequencies. We illustrate this structure with a linear function, $\tilde{G}$, in the Figure. The values of the Grey cells are 0.







Spatially Localized Spectral Conv:

- ✓ Spectral convolution is scale equivariant
- ✗ Global in nature – ill fit for vision tasks
- We propose Localized Spectral Convolution
  - ✓ Maintains equivariance
  - ✓ Captures local features

Non-Linearity and Pooling:

- For any function σ, its scale equivariant version, $\sigma_{eq}(\mathbf{x}) = \mathbf{y}$, can be defined as

$$\mathbf{Y}[k] = \mathcal{F}\Big(\sigma \circ \mathcal{F}^{-1}(\mathbf{X}[-|k|:|k|])\Big)[k]$$

where $\mathbf{X} = \mathcal{F}(\mathbf{x})$ and $\mathbf{Y} = \mathcal{F}(\mathbf{y})$

Classification and Scaling:

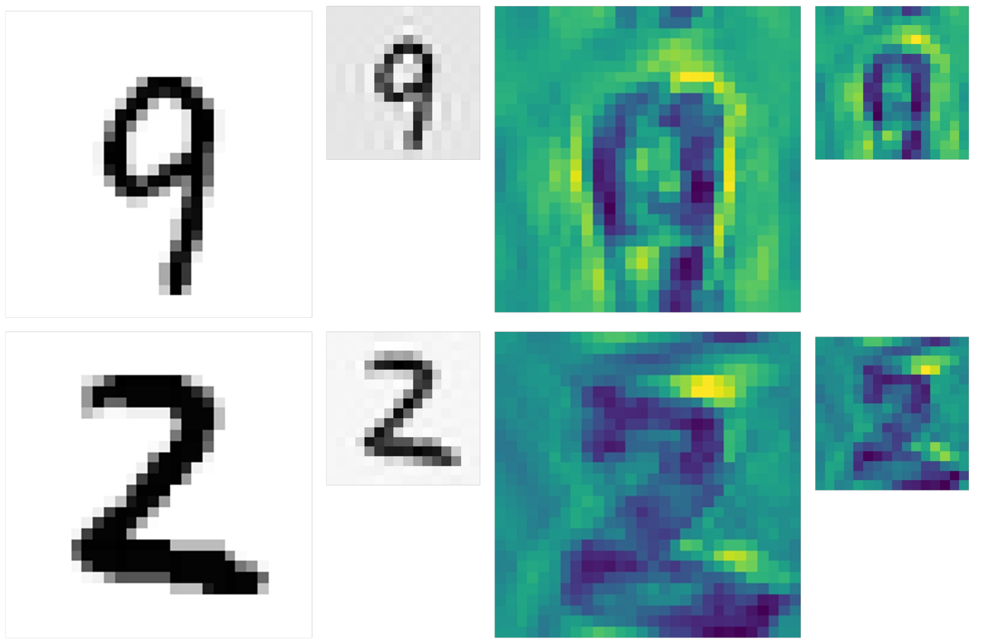
- Invariance is not desirable for scaling
- Classification loss should not increase at a higher scale
- We propose scale consistency loss

$$\mathcal{L}_{con} = \max\big(\mathcal{L}(\hat{\boldsymbol{y}}[k], y) - \mathcal{L}(\hat{\boldsymbol{y}}[k-1], y), 0\big).$$

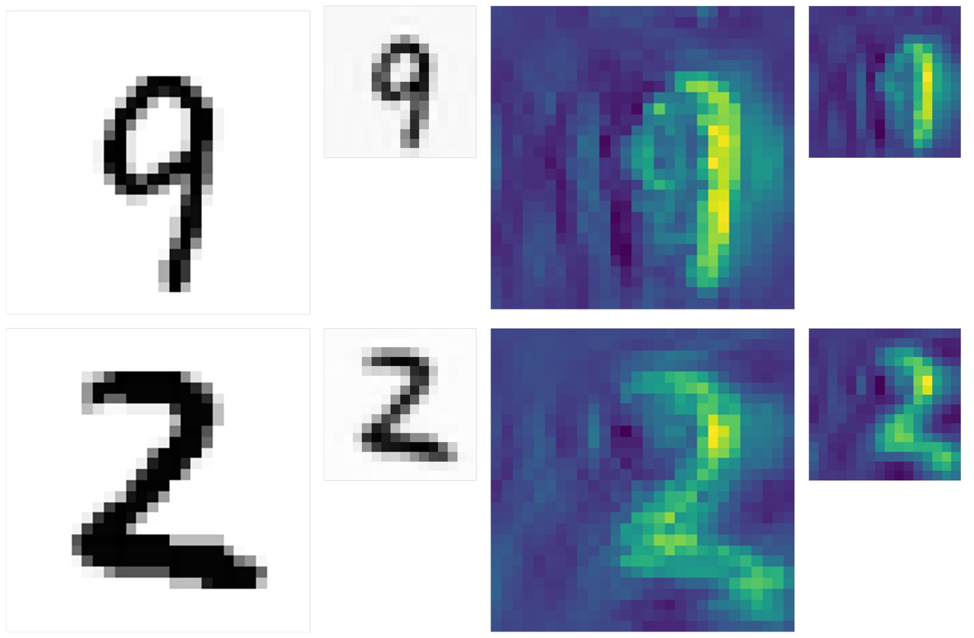
Here, $\hat{\boldsymbol{y}}[k]$ is the prediction of classifier $\mathcal{M}$ at k^{th} scale.

RESULTS

Qualitative Results: Perfect scale equivalence with non-linearity and pooling (Left). Robust to non-ideal down-sampling (Right).



Ideal Downsampling



Non-Ideal Downsampling

Quantitative Results:

Results on MNIST-scale

Models	Acc.↑	Scale-Con.↑	Equi-Err.↓
CNN	0.9737	0.6621	-
Per Res. CNN	0.9388	0.0527	-
SESN	0.9791	0.6640	-
DSS	0.9731	0.6503	-
SI-CovNet	0.9797	0.6425	-
SS-CNN	0.9613	0.3105	-
DISCO	0.9856	0.5585	0.44
Fourier CNN	0.9713	0.2421	0.28
Ours	**0.9889**	**0.9716**	**0.00**

Results on MNIST-scale with missing Scales

Models	Acc.↑	Scale-Con.↑	Equi-Err.↓
CNN	0.9737	0.6621	-
Per Res. CNN	0.9388	0.0527	-
SESN	0.9791	0.6640	-
DSS	0.9731	0.6503	-
SI-CovNet	0.9797	0.6425	-
SS-CNN	0.9613	0.3105	-
DISCO	0.9856	0.5585	0.44
Fourier CNN	0.9713	0.2421	0.28
Ours	**0.9889**	**0.9716**	**0.00**

Data Efficiency

Models / # Samples	5000	2500	1000
CNN	0.9432	0.9389	0.8577
Per Res. CNN	0.9118	0.8392	0.5815
DISCO	0.9794	0.9665	0.9457
SESN	0.9638	0.9402	0.9207
SI-CovNet	0.9641	0.9437	0.9280
SS-CNN	0.9477	0.9259	0.9176
DSS	0.9654	0.9401	0.9281
Fourier CNN	0.9567	0.9419	0.8910
Ours	**0.9835**	**0.9767**	**0.9606**

Results on STL10-scale

Models	Acc.↑	Scale-Con.↑	Equi-Err.↓
Wide ResNet	0.5596	0.2916	0.16
SESN	0.5525	0.4166	0.04
DSS	0.5347	0.1979	0.02
SI-CovNet	0.5588	0.2187	0.03
SS-CNN	0.4788	0.1979	1.82
DISCO	0.4768	0.3541	0.06
Fourier CNN	0.5844	0.2812	0.19
Ours	**0.7332**	**0.6770**	**0.00**