Truly shift equivariant convolutional neural networks with adaptive polyphase upsampling

Anadi Chaman*, Ivan Dokmanić†

*Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign

†Department of Mathematics and Computer Science, University of Basel

2021 Asilomar Conference on Signals, Systems, and Computers

Introduction

Shift equivariance is a desirable property for image reconstruction tasks.

$$F(T_k(x)) = T_k(F(x))$$

Popular CNN architectures like U-Net are not shift equivariant due to downsampling (stride) layers.

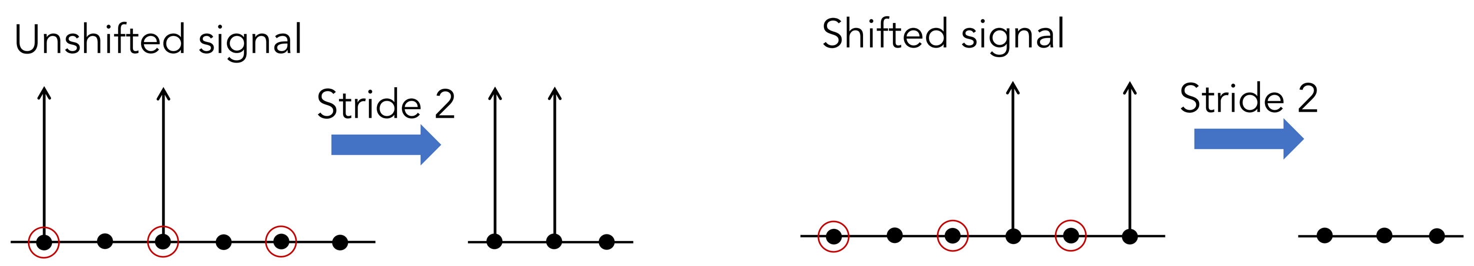
Existing solutions like data augmentation and anti-aliasing do not enable perfect equivariance.

- Gains in equivariance do not extend beyond training distribution.
- Stability to shifts improved on average but worst-case shifts can degrade performance.

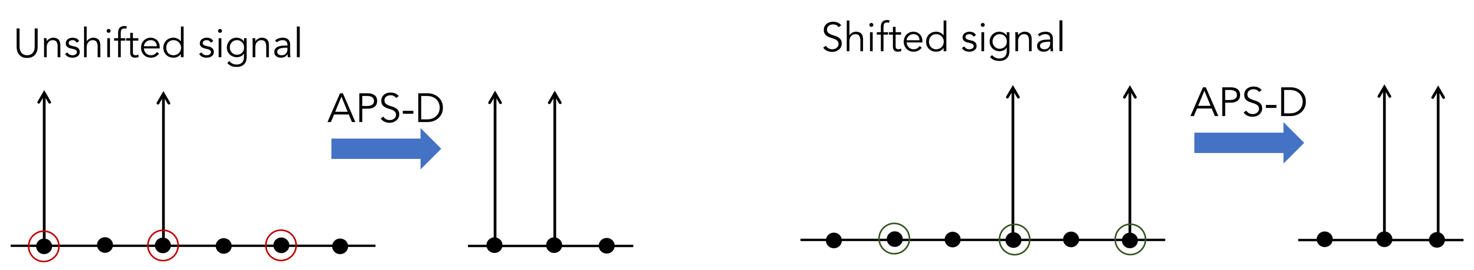
Our goal: Design a truly shift-equivariant CNN architecture for image-to-image regression problems without sacrificing performance.

Primer: Adaptive polyphase downsampling

Conventional downsampling is not robust to shifts.



Key idea: Choose the sampling grid adaptively.

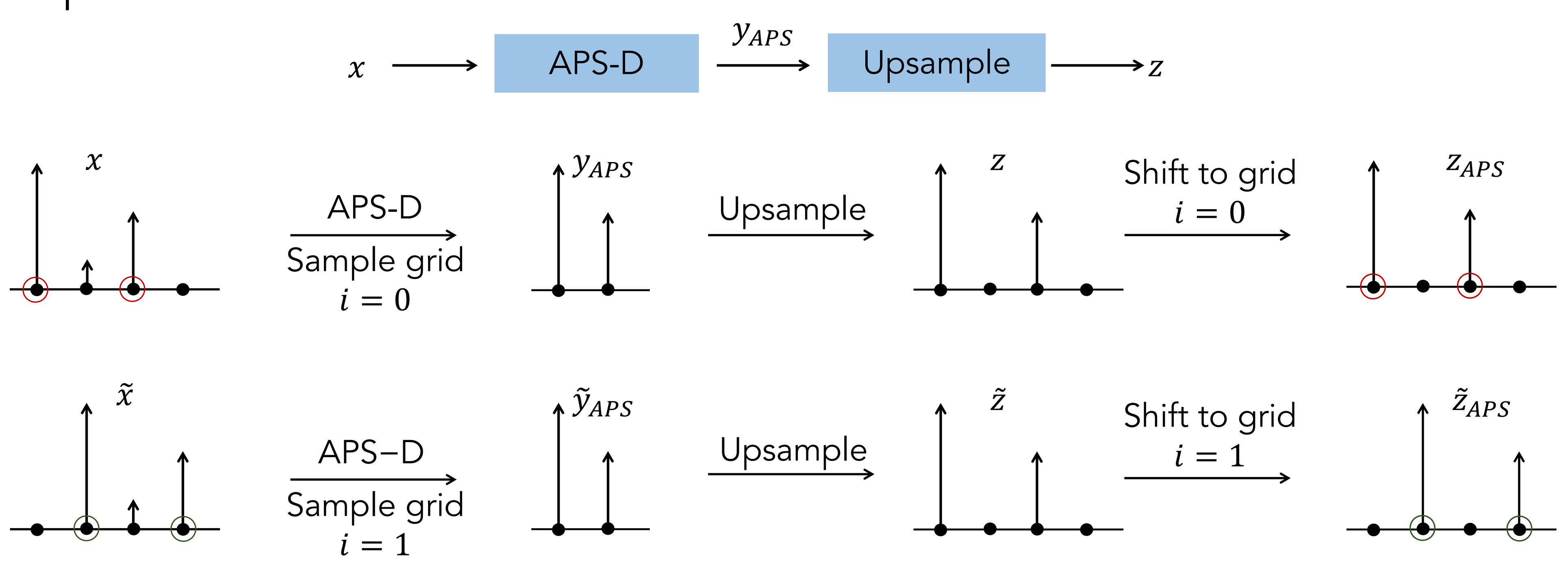


Adaptive polyphase downsampling chooses the sampling grid with the highest norm.

Shifting the input to APS-D shifts its subsampled output.

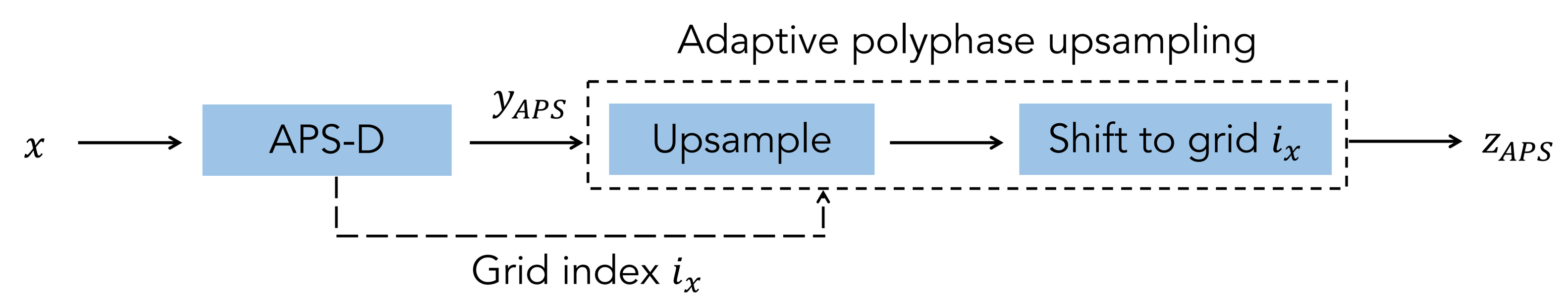
A case for adaptive upsampling

Replacing downsampling layers of a U-Net by APS-D is not sufficient for shift equivariance.



By upsampling the signal onto the grid used by APS-D, shift equivariance can be achieved.

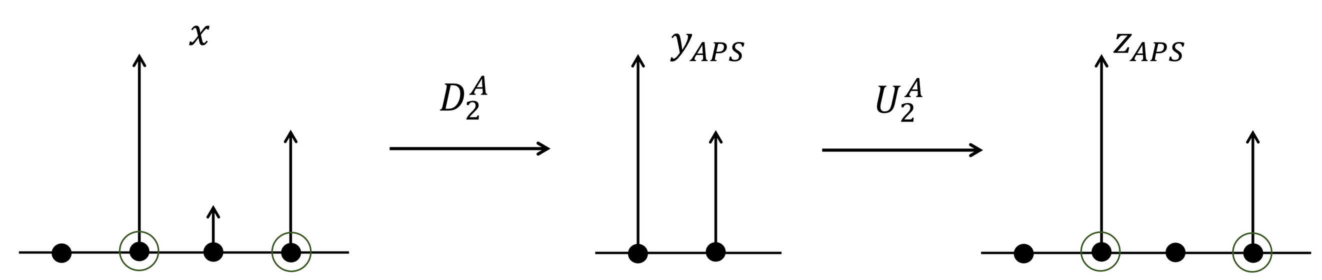
Adaptive polyphase upsampling



$$U_2^A(y_{APS}, i_x) = T_{i_x}(U_2(y_{APS}))$$

Proposition 1. *Let D_2^A and U_2^A denote APS-D and APS-U operators with stride 2. Then, $U_2^A \circ D_2^A$ is shift equivariant, i.e.*

$$U_2^A \circ D_2^A(T_k(x)) = T_k\left(U_2^A \circ D_2^A(x)\right), \qquad \forall\, k \in Z.$$



$U_2^A \circ D_2^A(x)$ zeros out all pixels in x, except those on the grid selected by APS-D.

Proposition 2. *A U-Net architecture with downsampling and upsampling layers replaced by APS-D and APS-U layers respectively is shift equivariant.*

Results

MRI reconstruction

U-Net variants with APS layers exhibit orders of magnitude superior equivariance compared to other methods.

Model	Equivariance metrics				Reconstruction metrics (unshifted)					
	NMSE		SSIM		NMSE		PSNR		SSIM	
	PD	PDFS	PD	PDFS	PD	PDFS	PD	PDFS	PD	PDFS
Baseline	0.0014	9.56e-4	0.9965	0.9975	0.016	0.053	33.83	29.92	0.8093	0.6301
LPF-3	4.97e-4	4.13e-4	0.9984	0.9988	**0.016**	**0.052**	**33.95**	**29.96**	**0.8125**	**0.6325**
Baseline + DA	1.37e-4	1.26e-4	0.9987	0.9990	0.016	0.053	33.58	29.86	0.8034	0.6272
APS	**1.21e-7**	**7.37e-15**	**1.0**	**1.0**	0.017	0.054	33.4	29.79	0.8013	0.6244
APS-3	**3.10e-7**	**1.36e-7**	**1.0**	**1.0**	**0.016**	**0.052**	**33.95**	**29.96**	**0.8124**	**0.6325**

Tab 1. Results obtained with different variants of U-Net on fastMRI validation set.

Out-of-distribution results

Networks with APS layers exhibit perfect shift equivariance even on out-of-distribution images.

Model	Baseline	LPF-2	Baseline + DA	APS	APS-2
NMSE	8.67e-3	4.65e-3	1.87e-3	**3.09e-14**	**3.20e-14**
SSIM	0.9722	0.9816	0.9822	**1.0**	**1.0**

Tab 2. Equivariance metrics for networks trained on fastMRI training set but evaluated on ImageNet validation set.

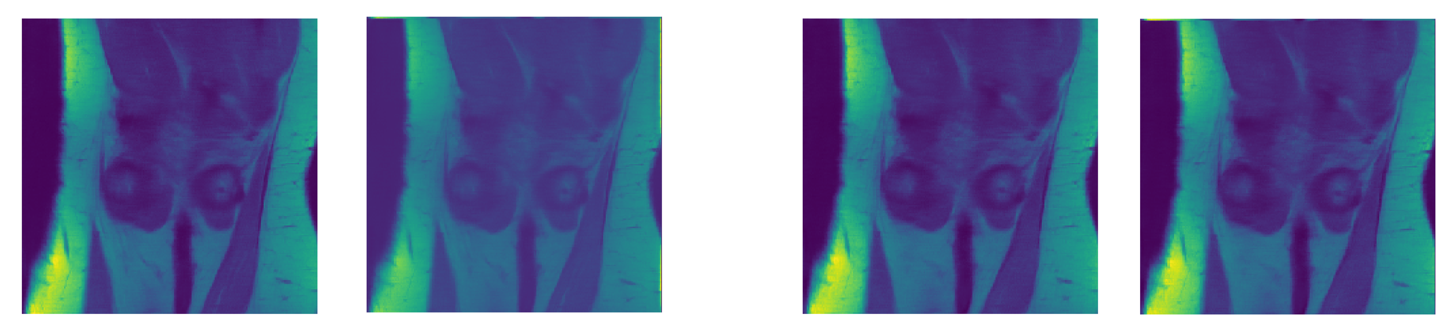
Decline in PSNR with shifts in input

With APS, U-Net reconstructions are significantly more stable to shifts than baseline.

Model	Baseline	LPF-2	Baseline + DA	APS	APS-2
(ΔPSNR)	4.03	2.58	0.037	**8.6e-4**	**5.87e-4**

Tab 4. Decline in PSNR of MRI reconstructions caused by randomly shifting the images in fastMRI validation set.

Baseline U-Net reconstructions — U-Net-APS reconstructions



Original output — Output for 3-pixel shifted input — Original output — Output for 3-pixel shifted input

Conclusion

- CNNs lack shift equivariance due to downsampling (stride).
- We propose adaptive polyphase upsampling to restore shift equivariance in symmetric encoder-decoder CNN architectures.
- APS provides SOTA in- and out-of-distribution equivariance performance without sacrificing reconstruction quality.

Code available at https://github.com/achaman2/truly_shift_invariant_cnns.