Making Convolutional Networks Shift-Invariant Again

Richard Zhang
Adobe Research
Example classifications

P(correct class) = 86.7

P(correct class) = 69.2
Deep Networks are **not** Shift-Invariant

\[
P(\text{correct class}) = 46.3 \quad \text{vs.} \quad 18.0
\]
Deep Networks are **not** Shift-Invariant


Why is shift-invariance lost?
Why is shift-invariance lost?

“Convolutions are shift-equivariant”
Why is shift-invariance lost?

“**Convolutions** are **shift-equivariant**”

“**Pooling** builds up **shift-invariance**”
Why is shift-invariance lost?

“**Convolutions** are shift-equivariant”

“**Pooling** builds up shift-invariance”

…but **striding** ignores Nyquist sampling theorem and **aliases**
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling
Re-examining Max-Pooling

Max-pooling breaks shift-equivariance
Shift-equivariance in VGG

• CIFAR

• VGG network
  • 5 max-pools
Shift-equivariance in VGG

• CIFAR

• VGG network
  • 5 max-pools

• Test shift-equivariance condition
  • \( \text{dist}(F(\text{Shift}_{\Delta h,\Delta w}(X)), \text{Shift}_{\Delta h,\Delta w}(F(X))) \)
Shift-equivariance, per layer

Perfect shift-equivariance, per layer
Shift-equivariance, per layer

Convolution is shift-equivariant

Large deviation from shift-eq.

Convolution is shift-equivariant
Shift-equivariance, per layer

Perfect shift-equivariance, per layer

Large deviation from shift-eq.

pool1

Δw

Δh

softmax
classifier

pool5

conv5

pool4

conv4

pool3

conv3

pool2

conv2

pool1

conv1

pixels
Perfect shift \(-eq.\)

Large deviation from shift-\(eq.\)

Shift-equivariance, per layer

Pooling breaks shift-equivariance

softmax
classifier
pool5
conv5
pool4
conv4
pool3
conv3
pool2
conv2
pool1
conv1
pixels
Shift-equivariance, per layer

- pool2
- Large deviation from shift-eq.
- Perfect shift-eq.
- softmax classifier
- pool5
- conv5
- pool4
- conv4
- pool3
- conv3
- pool2
- conv2
- pool1
- conv1
- pixels
Shift-equivariance, per layer

Large deviation from shift-eq.

Perfect shift-eq.
Shift-equivariance, per layer

Perfect shift-eq.

Large deviation from shift-eq.
Shift-equivariance, per layer

Nyquist theorem ignored when pooling; **aliasing** breaks shift-equivariance

Large deviation from shift-eq.
Alternative downsampling methods

- Blur+subsample
  - Antialiasing in signal processing; image processing; graphics
Alternative downsampling methods

• Blur+subsample
  • Antialiasing in signal processing; image processing; graphics

• Max-pooling
  • Performs better in deep learning applications [Scherer 2010]
Alternative downsampling methods

• Blur+subsample
  • Antialiasing in signal processing; image processing; graphics

• Max-pooling
  • Performs better in deep learning applications [Scherer 2010]

Reconcile antialiasing with max-pooling
Baseline (MaxPool)

max()  

max()  

heavy aliasing
Baseline (MaxPool)

heavy aliasing
Baseline (MaxPool)

heavy aliasing

(1) Max (dense evaluation)
no aliasing
Baseline (MaxPool)

max()

max()

heavy aliasing

(1) Max (dense evaluation) + (2) Subsampling

max()

max()

no aliasing

heavy aliasing
Baseline (MaxPool) = (1) Max (dense evaluation) + (2) Subsampling

heavy aliasing

no aliasing

heavy aliasing
Baseline (MaxPool)

(1) Max (dense evaluation) + (2) Subsampling
heavy aliasing

no aliasing

(1) Max (dense evaluation) + (2) Anti-aliasing filter

no aliasing

no aliasing

Anti-aliased (MaxBlurPool)
Baseline (MaxPool)

Max (dense evaluation) + Subsampling

Anti-aliased (MaxBlurPool)

Max (dense evaluation) + Anti-aliasing filter + Subsampling

Evaluated together as “BlurPool”
Antialiasing any downsampling layer

- Max Pool
  - VGG, Alexnet
Antialiasing any downsampling layer

- Max Pool
  - VGG, Alexnet

- Strided Convolution
  - Resnet, MobileNetv2
Antialiasing any downsampling layer

• Max Pool
  • VGG, Alexnet

• Strided Convolution
  • Resnet, MobileNetv2

• Average Pool
  • DenseNet
ImageNet

Shift-invariance
ImageNet

Shift-invariance

Accuracy
Antialiasing also improves accuracy
Discussion

Striding aliases \((\text{stride}=2)\)
Discussion

Striding aliases ($\text{stride}=2$)

Add antialiasing filter
Discussion

Striding aliases ($\text{stride}=2$)

Add antialiasing filter
  + Improved shift-equivariance
Discussion

Striding aliases ($\text{stride}=2$)

Add antialiasing filter
  + Improved shift-equivariance
  + Improved accuracy
Discussion

Striding aliases \((stride=2)\)

Add antialiasing filter
  + Improved shift-equivariance
  + Improved accuracy

Additionally
  + Improved stability to other perturbations
Discussion

Striding aliases (stride=2)

Add antialiasing filter
  + Improved shift-equivariance
  + Improved accuracy

Additionally
  + Improved stability to other perturbations
  + Improved robustness
Discussion

Striding aliases ($\text{stride}=2$)

Add antialiasing filter
  + Improved shift-equivariance
  + Improved accuracy

Additionally
  + Improved stability to other perturbations
  + Improved robustness

Antialiasing code, pretrained models
  https://richzhang.github.io/antialiased-cnns/
Discussion

Striding aliases \((\text{stride}=2)\)

Add antialiasing filter
  + Improved shift-equivariance
  + Improved accuracy

Additionally
  + Improved stability to other perturbations
  + Improved robustness

Antialiasing code, pretrained models
https://richzhang.github.io/antialiased-cnns/

Thank you!