On the Shift Invariance of Max Pooling Feature Maps in CNNs

Hubert Leterme, *hubert.leterme@univ-grenoble-alpes.fr*

Université Grenoble Alpes, Laboratoire Jean Kuntzmann
Inria Grenoble Rhône-Alpes, Thoth project-team

Context

Desired feature extraction properties for image classification:

51 retain discriminant image components (class separation);
51 reduce intra-class variability.
Key property: stability *w.r.t.* input transformations.

Oscillating patterns very often observed in CNN kernels.

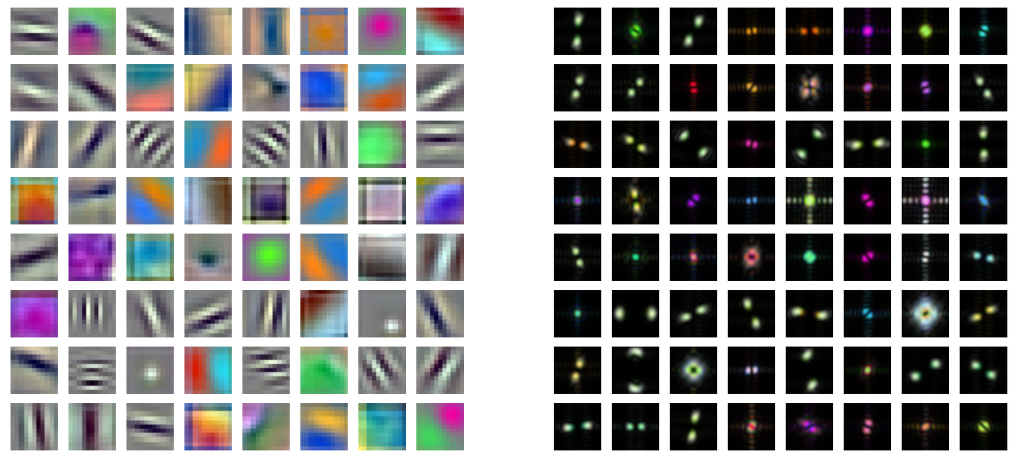
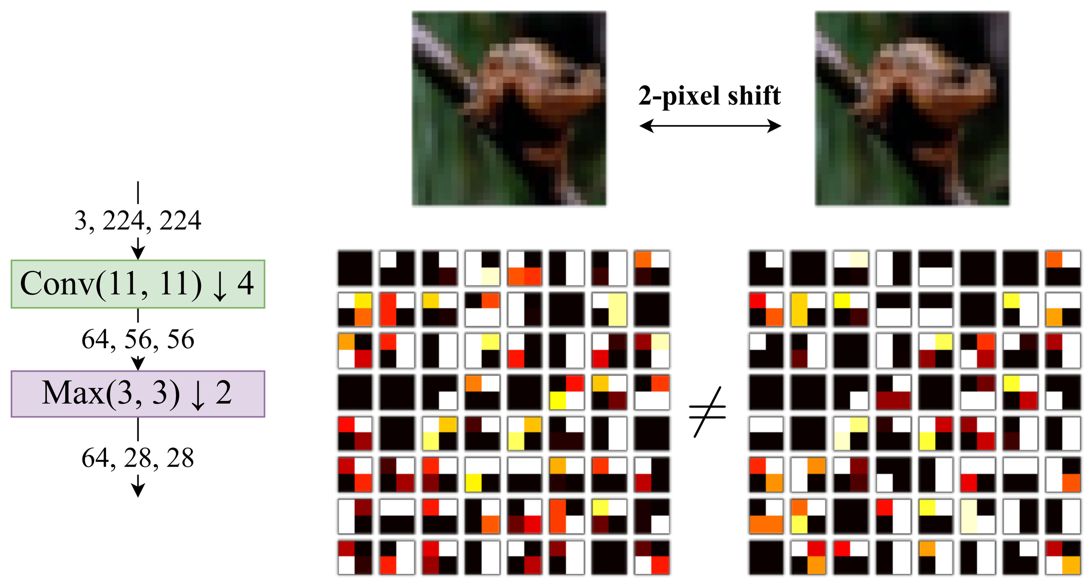


Fig. 1: AlexNet's first layer after training with ImageNet. Left: spatial domain. Right: Fourier domain.

Are CNN First Layers Shift-Invariant?



Objectives

1. Theoretical study: **establish a measure of shift invariance** for max pooling output feature maps.
2. Experimental study: **improve model stability** by using the tools introduced in the theoretical study.

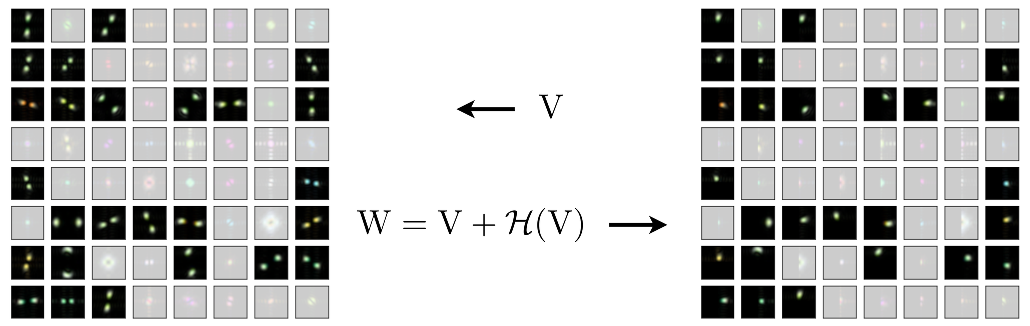
Related work

51 Preliminary result sketched by Waldspurger in the continuous case [Wal15]. **Does this extend to discrete subsampled convolutions?**

51 Invariance study done by Wiatowski and Bölcskei [WB18], in the continuous framework. **What about aliasing? What about max pooling?**

51 Antialiasing methods based on low-pass filtering [Zha19; Zou+20] led to improved accuracy, despite a loss of information. **Possibility to design an antialiasing method preserving high-frequency information?**

1a. Theoretical results



Operator in "standard" CNNs: $\mathbb{R}$Max

$\mathcal{R}_m : \mathrm{X} \mapsto ((\mathrm{X} \star \mathrm{V}) \downarrow m)$

$\mathcal{R}_m\mathrm{X}[n] := \max_{\|k\|_\infty \leq 1} \mathrm{Y}[2n+k]$, with $\mathrm{Y}[n] := (\mathrm{X} \star \mathrm{V})\,[mn]$.

Proxy operator: $\mathbb{C}$Mod

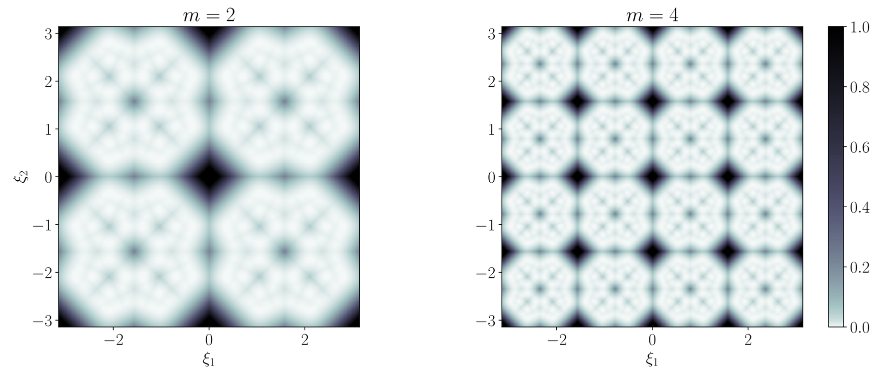
$\mathcal{C}_{2m} : \mathrm{X} \mapsto |(\mathrm{X} \star \mathrm{W}) \downarrow (2m)|$

$\mathcal{C}_{2m}\mathrm{X}[n] := |\mathrm{Z}[n]|$, with $\mathrm{Z}[n] := (\mathrm{X} \star \mathrm{W})\,[2mn]$.

The complex filter W is obtained by computing the **2D Hilbert transform** [HHB97] of the trained weight V, i.e., $\mathrm{W} := \mathrm{V} + i\mathcal{H}(\mathrm{V})$.

Theorem 1 (Stability of $\mathbb{C}$Mod). *If $\widehat{\mathrm{W}}$ is sufficiently localized, then $\mathcal{C}_{2m}(\mathcal{T}_q\mathrm{X}) \approx \mathcal{C}_{2m}\mathrm{X}$, where $\mathcal{T}_q$ denotes a translation operator with $q \in \mathbb{R}^2$.*

Theorem 2 (Similarity between $\mathbb{R}$Max and $\mathbb{C}$Mod). *If $\widehat{\mathrm{W}}$ is sufficiently localized, then $\|\mathcal{R}_m\mathrm{X} - \mathcal{C}_{2m}\mathrm{X}\|_2$ remains small,* ***except for some pathological frequencies****.*



Corollary (Stability of $\mathbb{R}$Max). *The shift invariance of $\mathbb{R}$Max* ***depends on the characteristic frequency*** *of W, as well as its Fourier resolution.*

1b. Shift invariance of max pooling outputs

Experiments conducted **50K images from ImageNet** (224×224 pixels), decomposed acc. to the **dual-tree wavelet packet transform (DT-$\mathbb{C}$WPT)** [BS08] with J levels of decomposition, s.t. $m = 2^{J-1}$ (subsampling factor in $\mathbb{R}$Max).

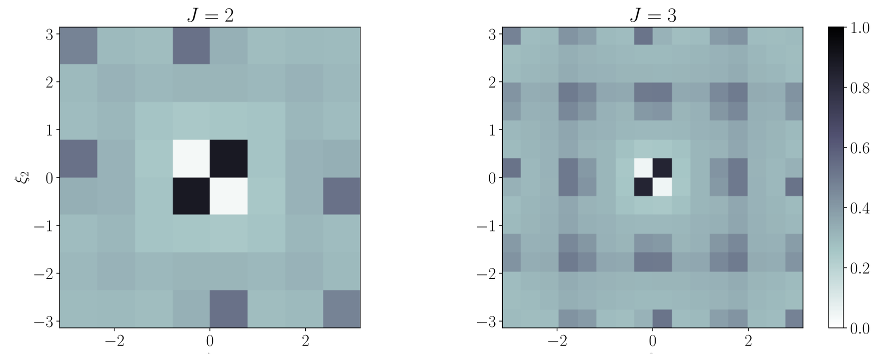


Fig. 5: Mean discrepancy between $\mathbb{R}$Max and $\mathbb{C}$Mod outputs.

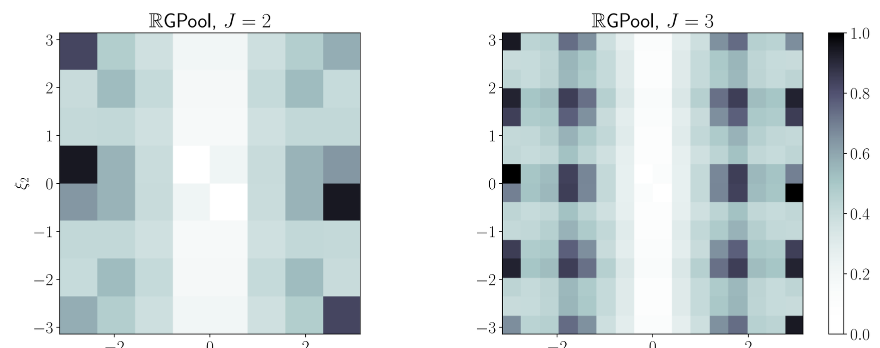
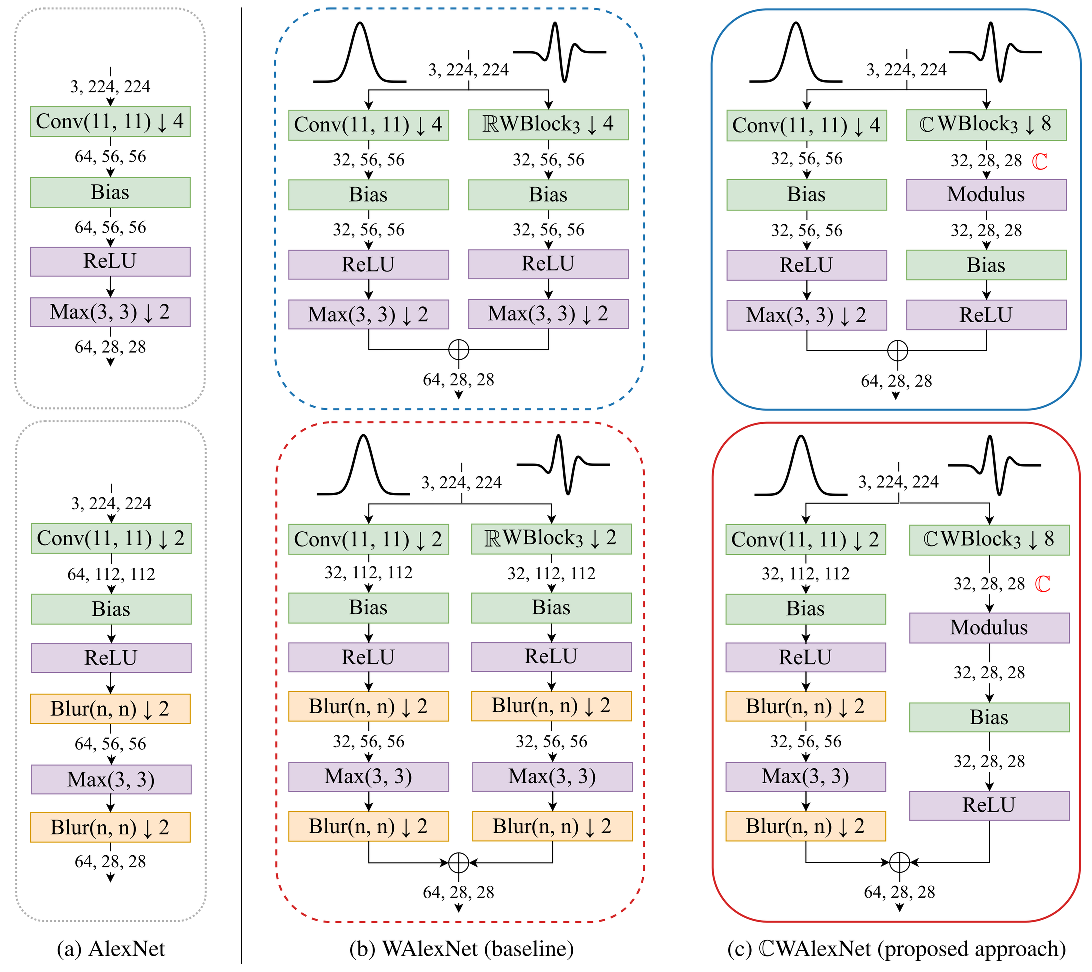


Fig. 6: Shift stability of $\mathbb{R}$Max outputs.

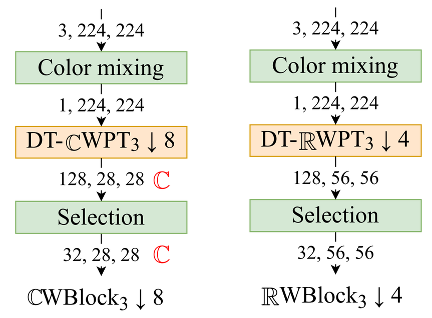
Each filter is represented as a pixel **localized around its charact. frequency**.

2a. Design of wavelet-based twin CNNs

51 Antialiasing principle based on Theorems 1 and 2: **replace $\mathbb{R}$Max by $\mathbb{C}$Mod in existing CNN architectures**.
55 **Only for high-frequency filters** with well-defined orientations **unpredictable in standard CNNs**.



	WAlexNet	WResNet
J (decomposition depth)	3	2
K_{dt} (dual-tree filters)	128	32
K'_{dt} (manually selected filters)	94	16
L_{low}, L_{high} (output channels)	32, 32	40, 24



2b. Evaluation scores

Model	One-crop top-1	One-crop top-5	Ten-crops top-1	Ten-crops top-5	Shifts mFR
	AlexNet				
Standard	*45.3*	*22.2*	*41.3*	*19.3*	*100.0*
$\mathbb{R}$Max^a	44.9	21.8	40.8	19.0	101.4
$\mathbb{C}$Mod^a	**44.3**	**21.3**	**40.2**	**18.5**	**88.0**
Blur	*44.8*	*22.0*	*41.1*	*19.1*	*58.1*
Blur$\mathbb{R}$Max^a + Blur	44.6	21.9	40.6	19.0	**59.2**
$\mathbb{C}$Mod^a + Blur	**43.6**	**20.9**	**39.5**	**17.9**	71.0
	ResNet-34				
Standard	*27.6*	*9.2*	*24.8*	*7.7*	*78.1*
$\mathbb{R}$Max^a	27.4	9.2	24.7	7.6	77.2
$\mathbb{C}$Mod^a	**27.2**	**9.0**	**24.4**	**7.4**	**73.1**
Blur	*26.5*	*8.7*	*24.1*	*7.3*	*60.3*
Blur$\mathbb{R}$Max^a + Blur	**26.6**	8.7	24.3	**7.3**	62.7
$\mathbb{C}$Mod^a + Blur	**26.6**	**8.6**	**24.0**	**7.3**	**61.5**
ABlur	*26.1*	*8.3*	*23.5*	*7.0*	*60.8*
ABlur$\mathbb{R}$Max^a + ABlur	**26.0**	**8.2**	**23.6**	**6.9**	**62.1**
$\mathbb{C}$Mod^a + ABlur	26.1	**8.2**	23.7	7.0	63.1

Fig. 9: Evaluation scores on ImageNet

Model	ResNet-18 1crp	ResNet-18 10crp	ResNet-18 shft	ResNet-34 1crp	ResNet-34 10crps	ResNet-34 shft
Standard	*14.9*	*10.8*	*100.0*	*15.2*	*10.9*	*100.3*
$\mathbb{R}$Max^a	14.2	10.3	92.4	14.5	10.5	99.2
$\mathbb{C}$Mod^a	**13.8**	**9.6**	**88.8**	**12.9**	**9.2**	**93.0**
Blur	*14.2*	*10.4*	*87.7*	*15.7*	*11.6*	*88.2*
Blur$\mathbb{R}$Max^a + Blur	13.1	9.7	**84.6**	13.2	9.9	85.6
$\mathbb{C}$Mod^a + Blur	**12.3**	**8.9**	85.7	**12.4**	**9.1**	**83.7**
ABlur	*14.6*	*11.0*	*90.9*	*16.3*	*12.8*	*91.9*
ABlur$\mathbb{R}$Max^a + ABlur	14.5	11.0	86.5	14.0	10.4	93.3
$\mathbb{C}$Mod^a + ABlur	**12.8**	**9.7**	**81.7**	**12.8**	**9.2**	**86.6**

Fig. 10: Evaluation scores on CIFAR-10

2c. Training and evaluation curves

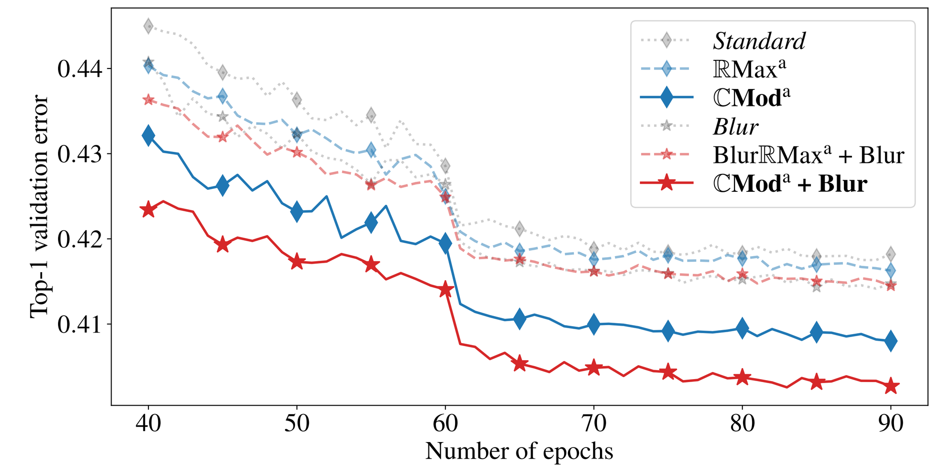


Fig. 11: Top-1 validation error on ImageNet, AlexNet-based models.

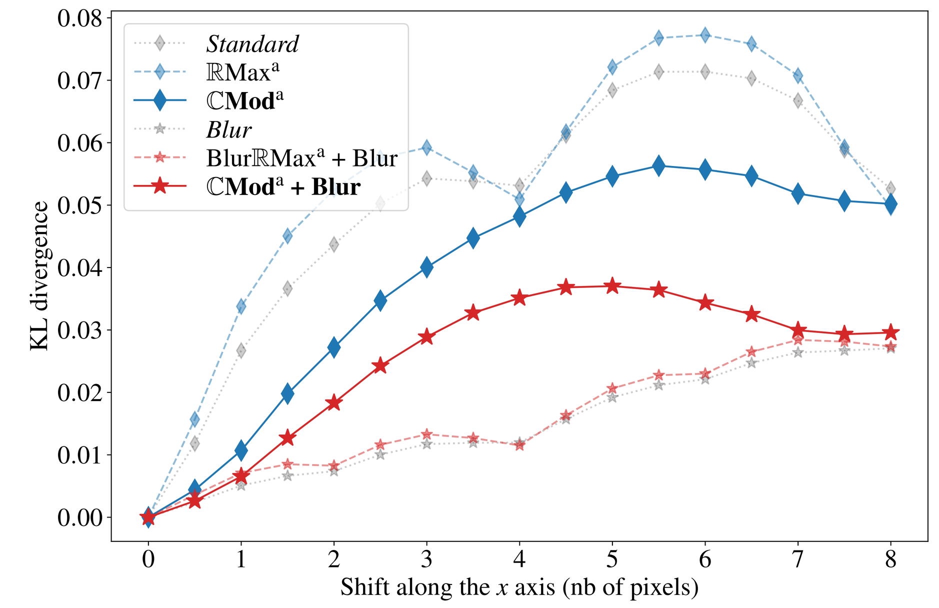


Fig. 12: Shift invariance of AlexNet-base models

References and acknowledgments

[BS08] I. Bayram and I. W. Selesnick. "On the Dual-Tree Complex Wavelet Packet and M-Band Transforms". In: *IEEE Trans. Signal Processing* 56.6 (2008).

[HHB97] J. Havlicek et al. "The Analytic Image". In: *Proc. ICIP*. 1997.

[Wal15] I. Waldspurger. "Wavelet Transform Modulus : Phase Retrieval and Scattering". Doctoral Thesis. ENS, Paris, 2015.

[WB18] T. Wiatowski and H. Bölcskei. "A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction". In: *IEEE Trans. Information Theory* 64.3 (2018).

[Zha19] R. Zhang. "Making Convolutional Networks Shift-Invariant Again". In: *ICML*. 2019.

[Zou+20] X. Zou et al. "Delving Deeper into Anti-aliasing in ConvNets". In: *BMVC*. 2020.

Submitted papers:

[1] H. Leterme, K. Polisano, V. Perrier, and K. Alahari, "On the Shift Invariance of Max Pooling Feature Maps in Convolutional Neural Networks", arXiv:2209.11740, Sep. 2022. Under review.

[2] H. Leterme, K. Polisano, V. Perrier, and K. Alahari, "From CNNs to Shift-Invariant Twin Wavelet Models", Nov. 2022. Under review.

This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir, as well as the ANR grant AVENUE (ANR-18-CE23-0011).

LaTeX TikZposter