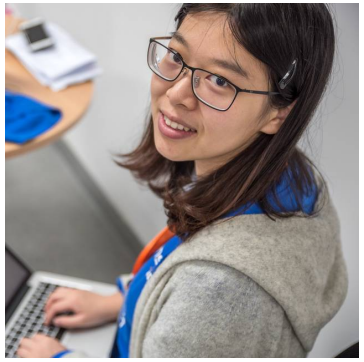
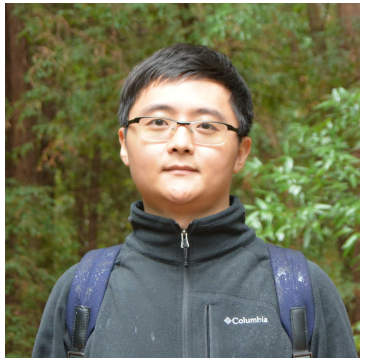
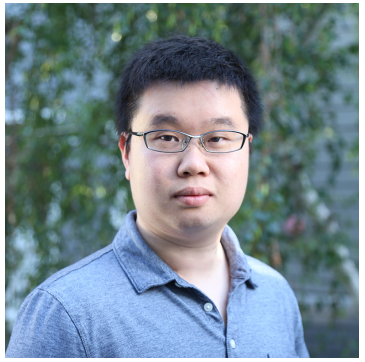
Delving Deeper into Anti-aliasing in ConvNets



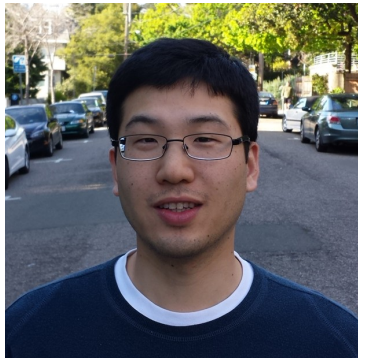
Xueyan Zou



Fanyi Xiao



Zhiding Yu

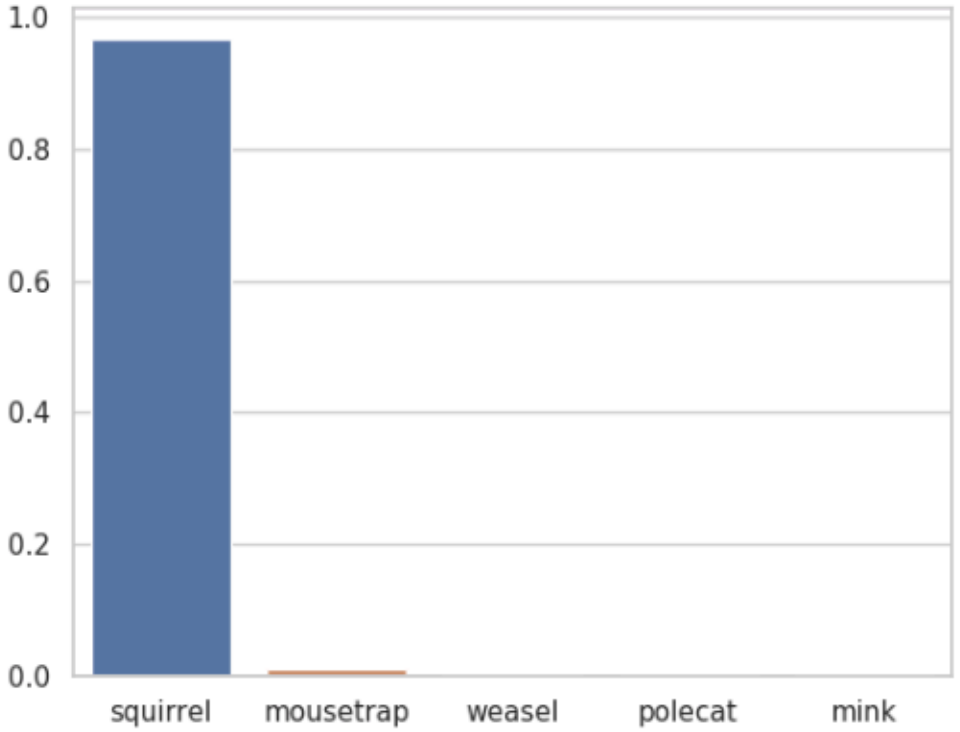


Yong Jae Lee

UCDAVIS

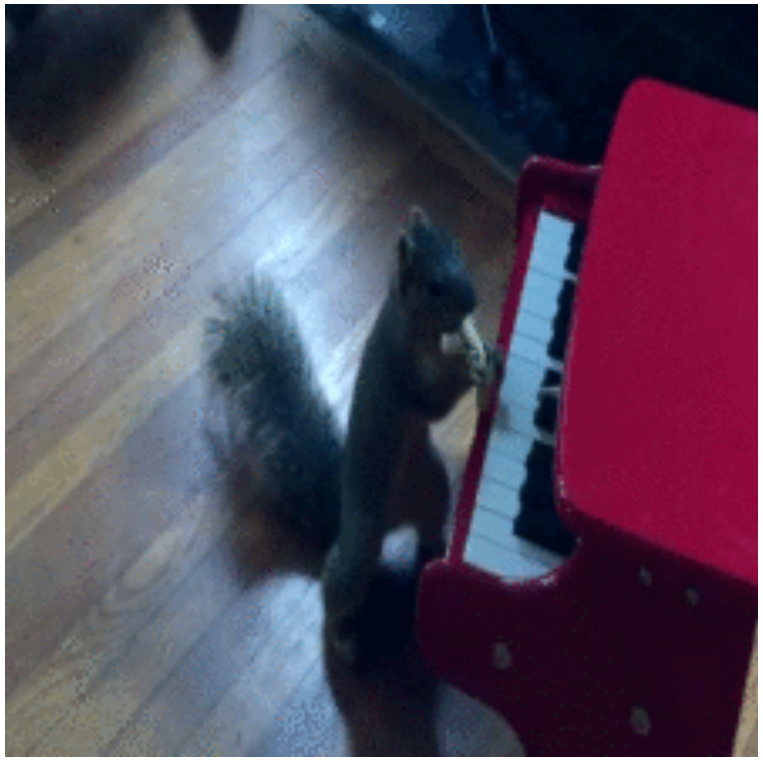
NVIDIA

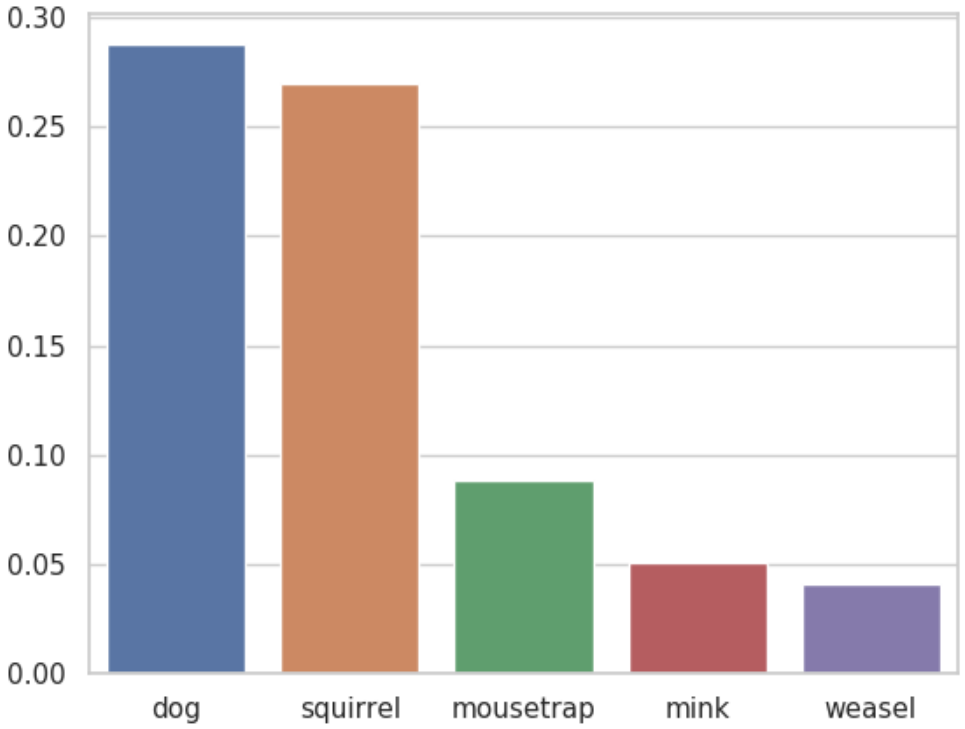
Shift Inconsistency in Image Classification



Classifier could correctly classify the image as squirrel

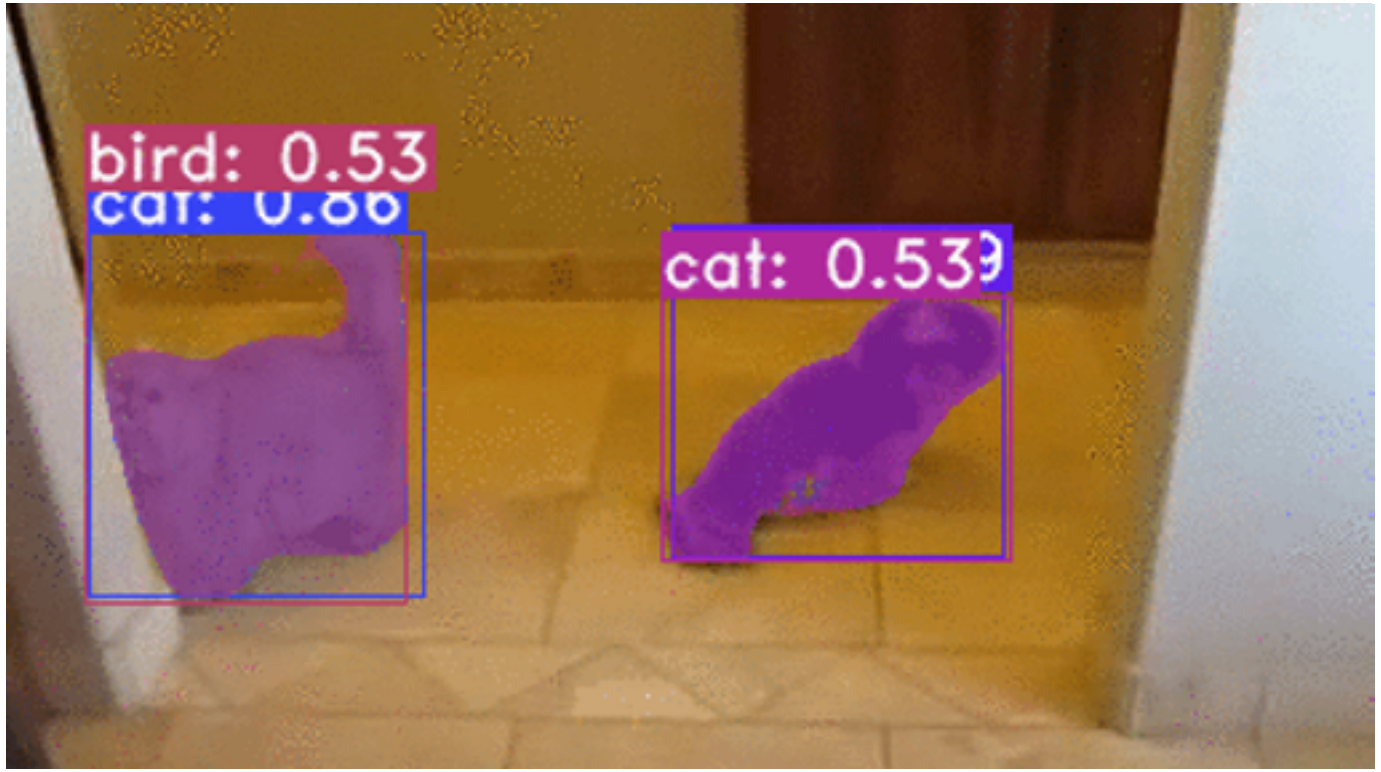
Shift Inconsistency in Image Classification





With a tiny shift, output prediction changes dramatically

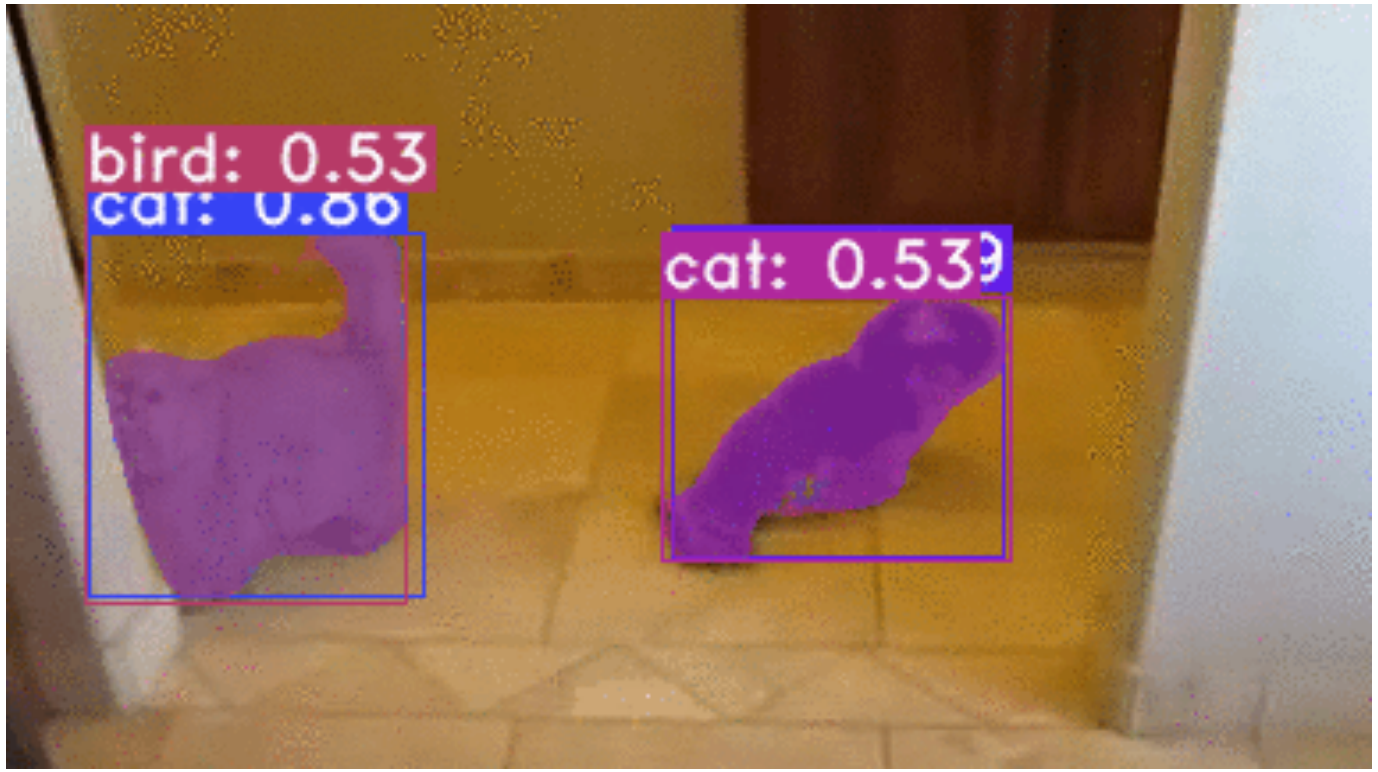
Shift Inconsistency in Instance Segmentation



This phenomenon holds for instance segmentation

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*
[2] Do Image Classifiers Generalize Across Time? Shankar et al. *arXiv 2019*

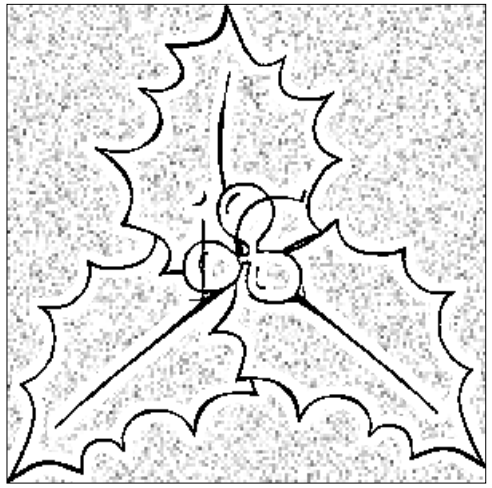
Shift Inconsistency in Instance Segmentation



Classifier hovers the decision between cat and dog

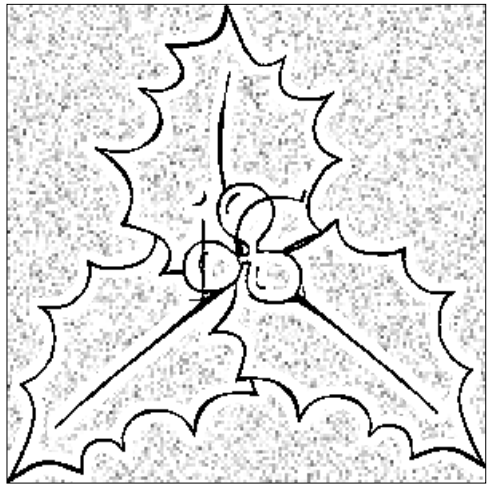
[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*
[2] Do Image Classifiers Generalize Across Time? Shankar et al. *arXiv 2019*

Aliasing



(a) Input

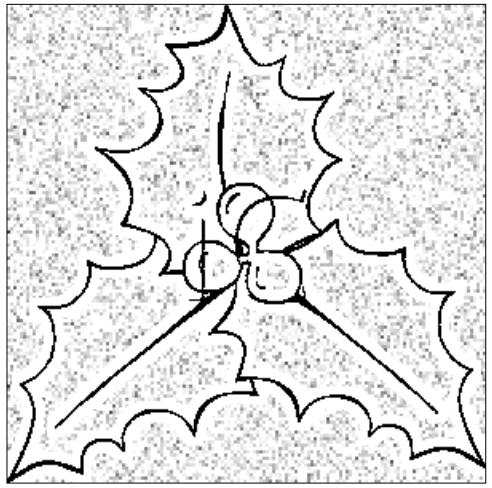
Aliasing



(a) Input

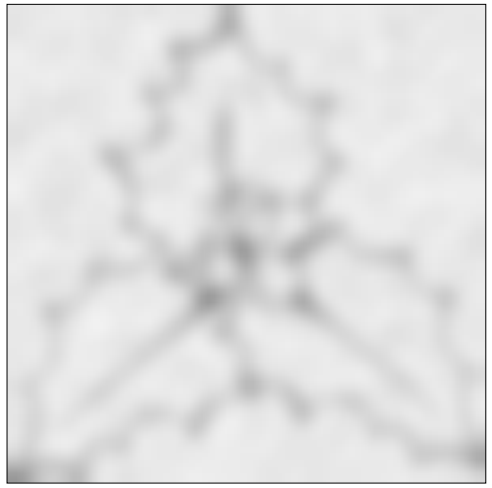
(b) 4x Down

Anti-aliasing



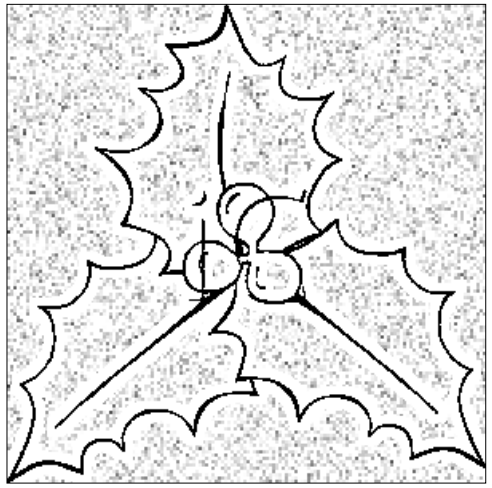
(a) Input

(b) 4x Down



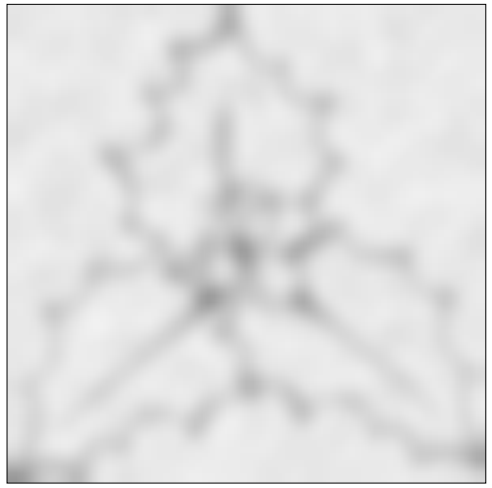
(c) Gaussian+4x Down

Adaptive Anti-aliasing

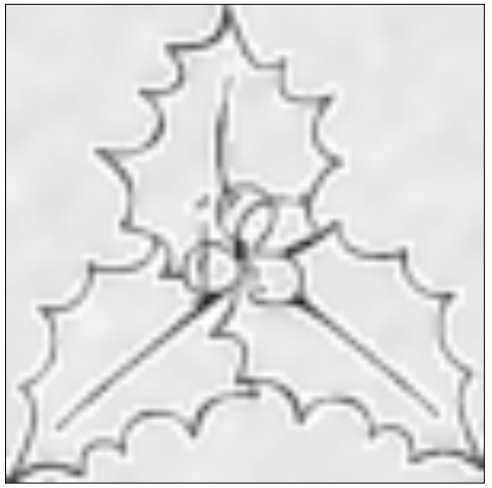


(a) Input

(b) 4x Down

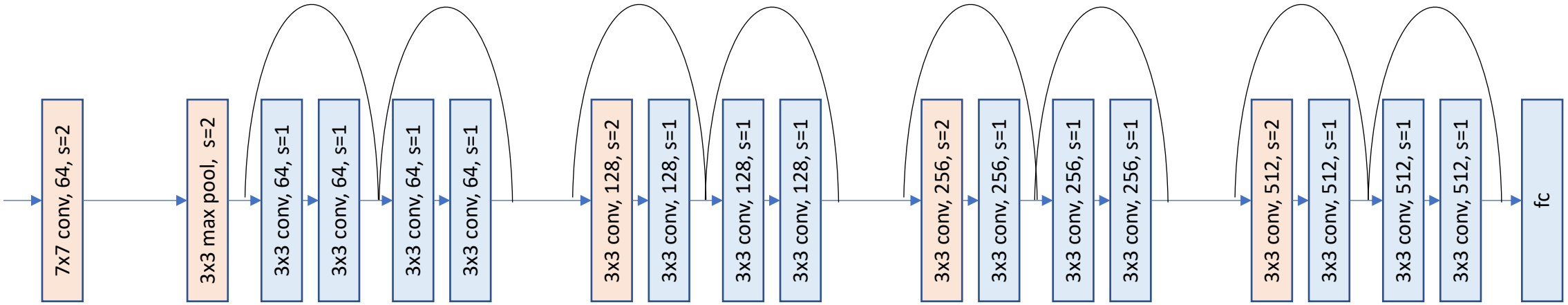


(c) Gaussian+4x Down



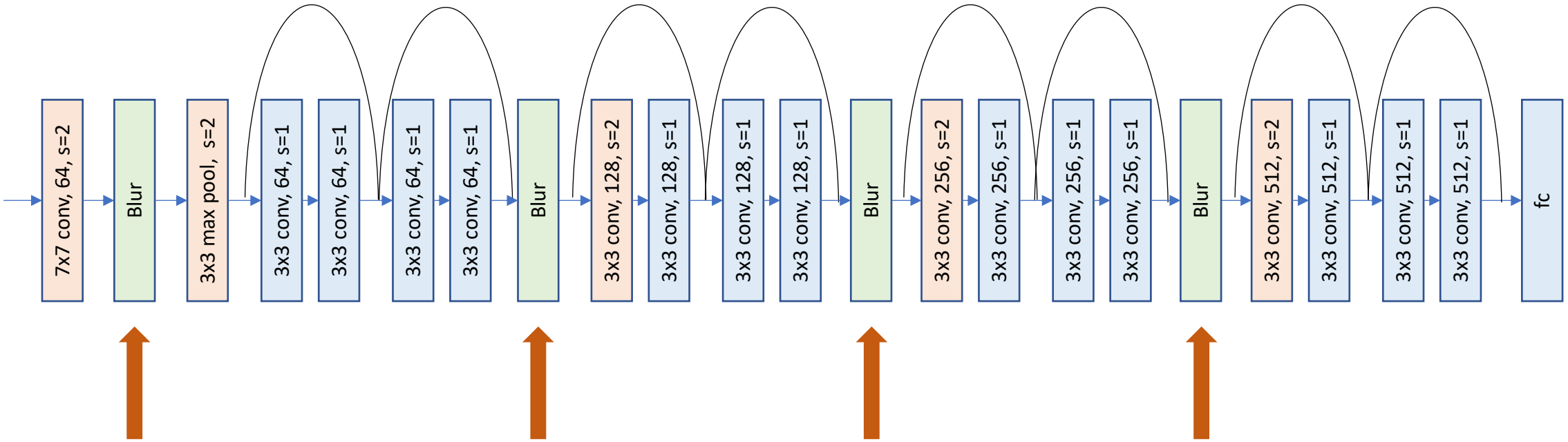
(d) Adaptive+4x Down

Anti-aliasing in ConvNets: ResNet 18



[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Anti-aliasing in ConvNets: ResNet 18



[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

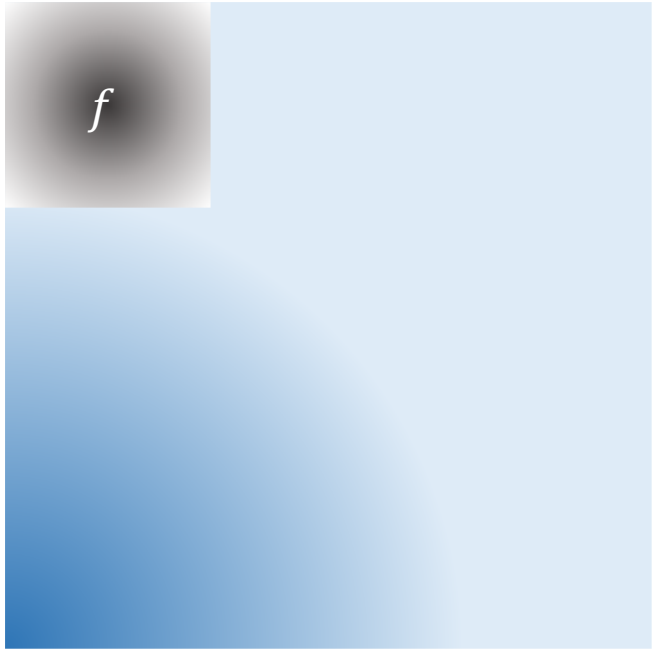
Fixed vs Spatially-Adaptive Low-pass filter

(a) Fixed Lowpass Filter [1]

(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-Adaptive Low-pass filter

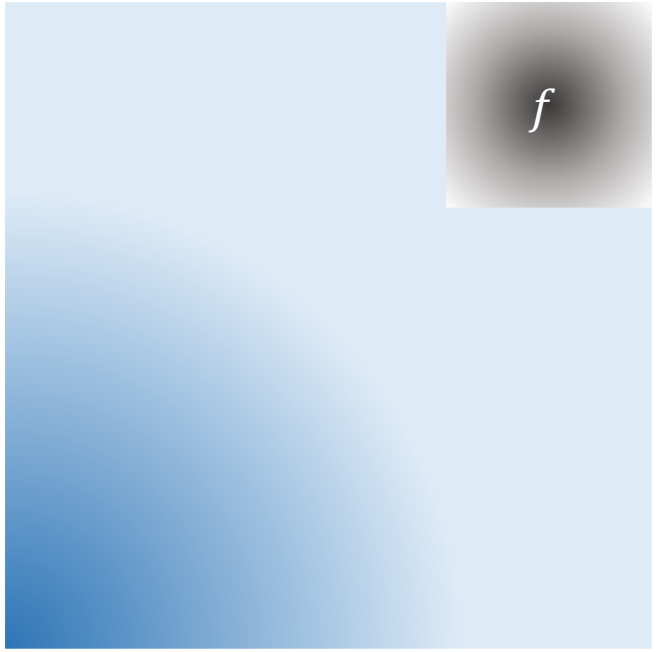


(a) Fixed Lowpass Filter [1]

(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-Adaptive Low-pass filter

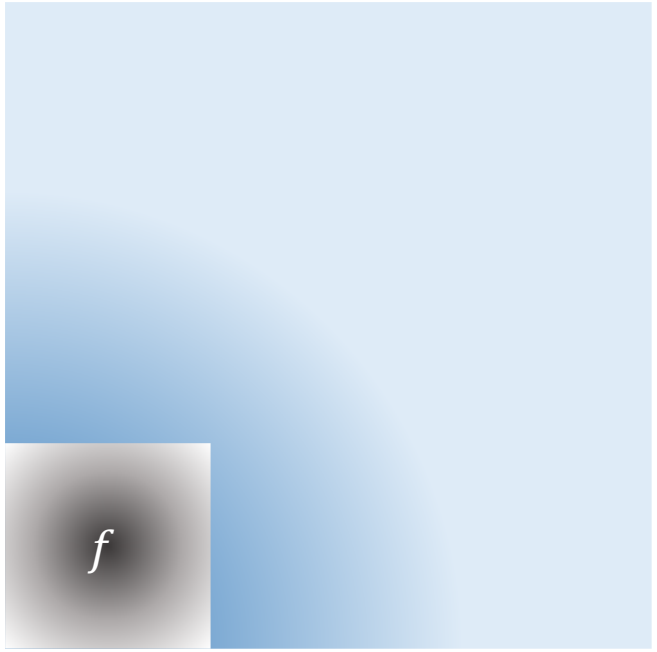


(a) Fixed Lowpass Filter [1]

(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-Adaptive Low-pass filter

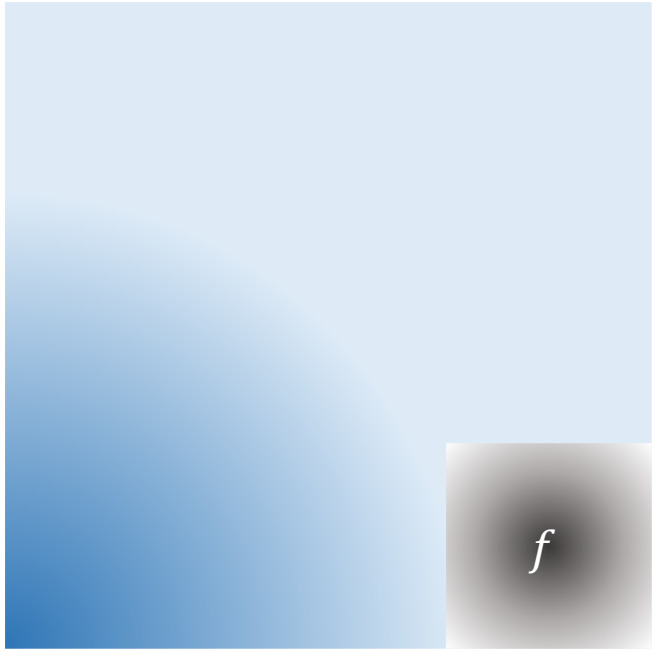


(a) Fixed Lowpass Filter [1]

(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-Adaptive Low-pass filter

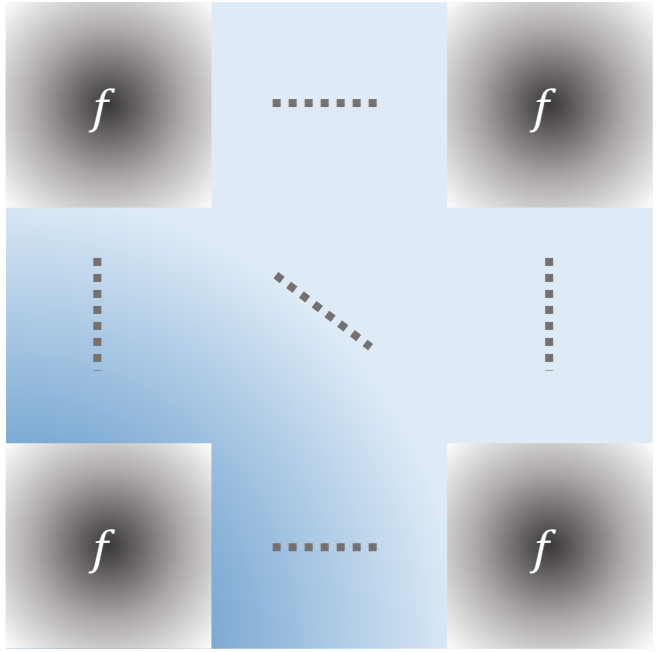


(a) Fixed Lowpass Filter [1]

(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Low-pass Filter vs Spatial Adaptive Low-pass filter

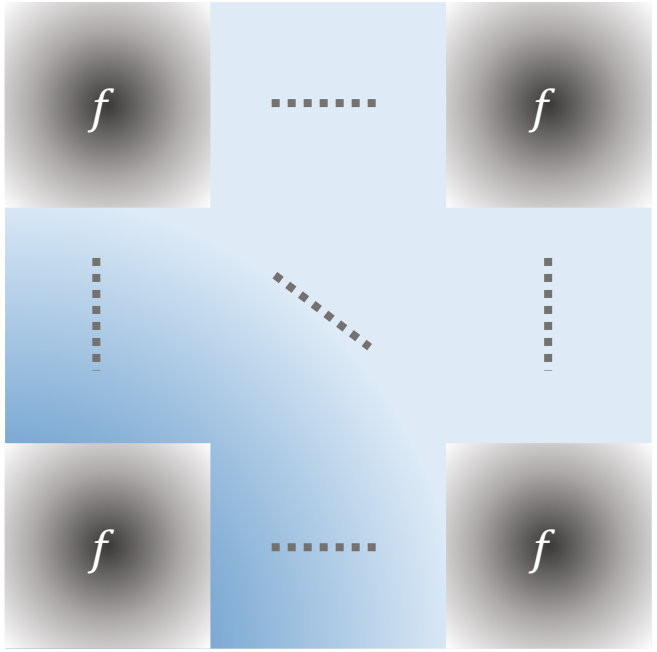


(a) Fixed Lowpass Filter [1]

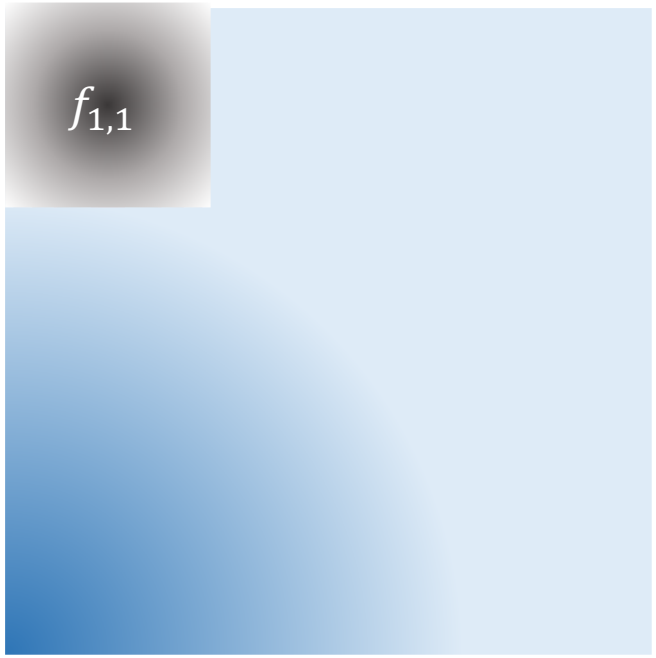
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-adaptive Low-pass filter



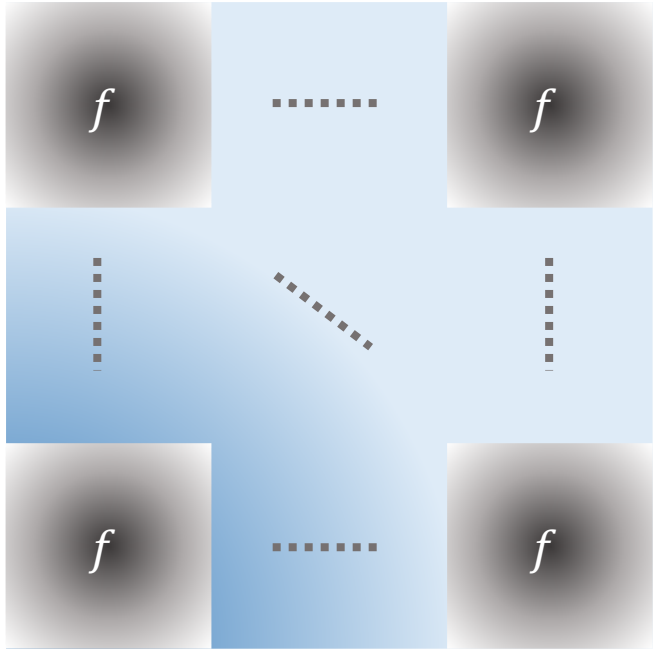
(a) Fixed Lowpass Filter [1]



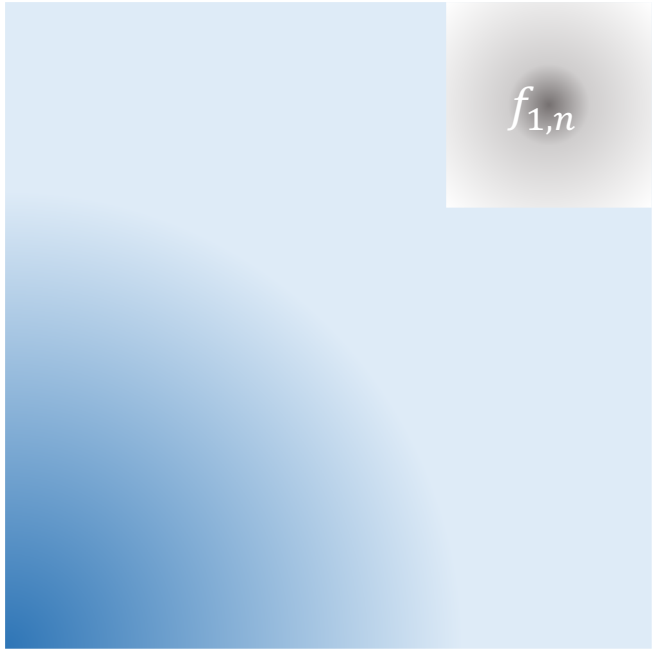
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-adaptive Low-pass filter



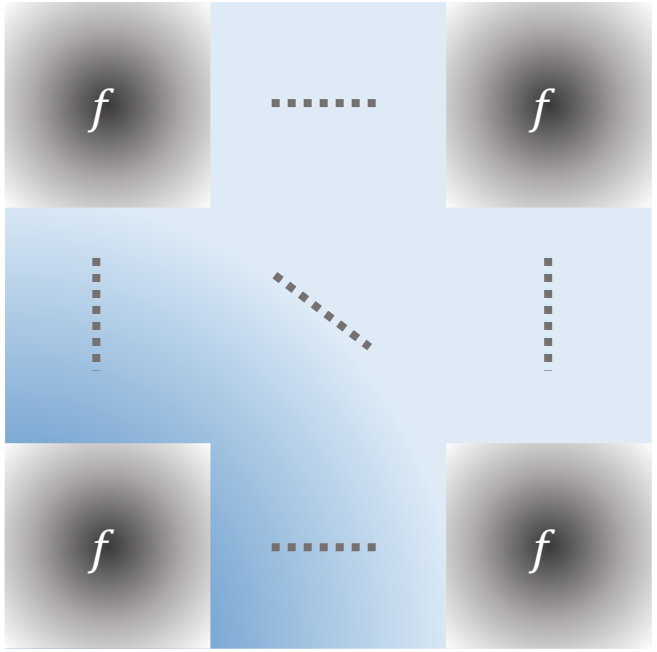
(a) Fixed Lowpass Filter [1]



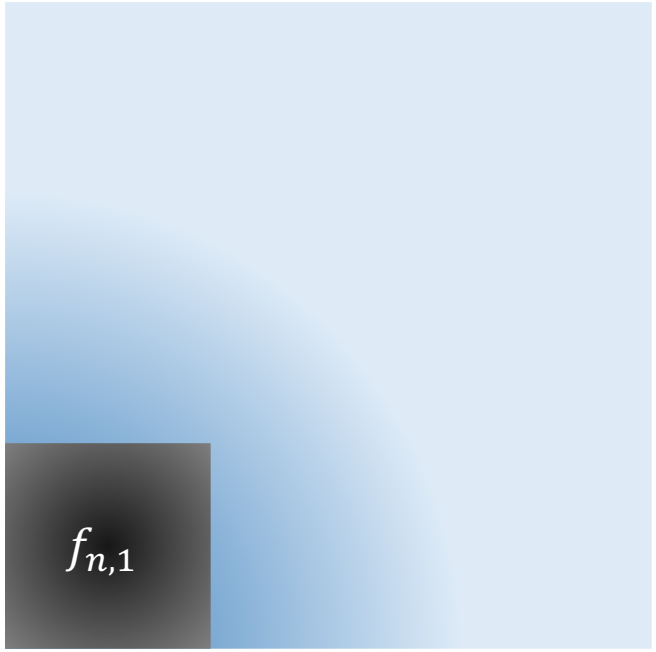
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-adaptive Low-pass filter



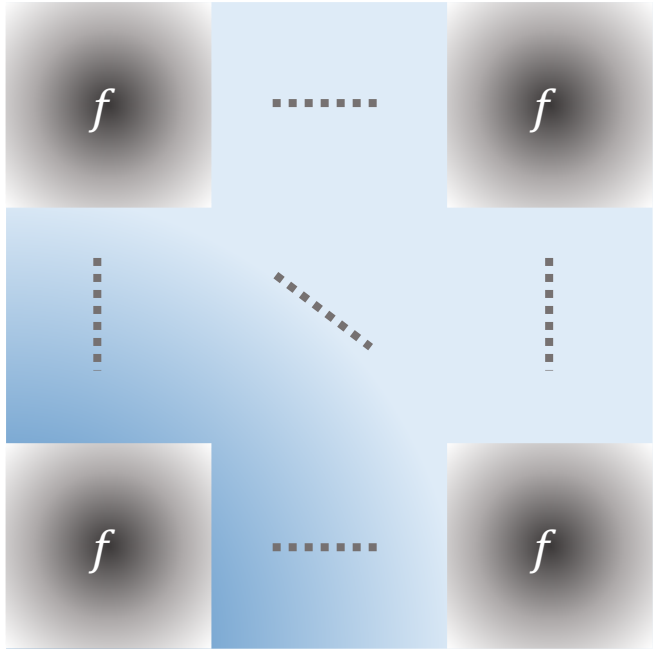
(a) Fixed Lowpass Filter [1]



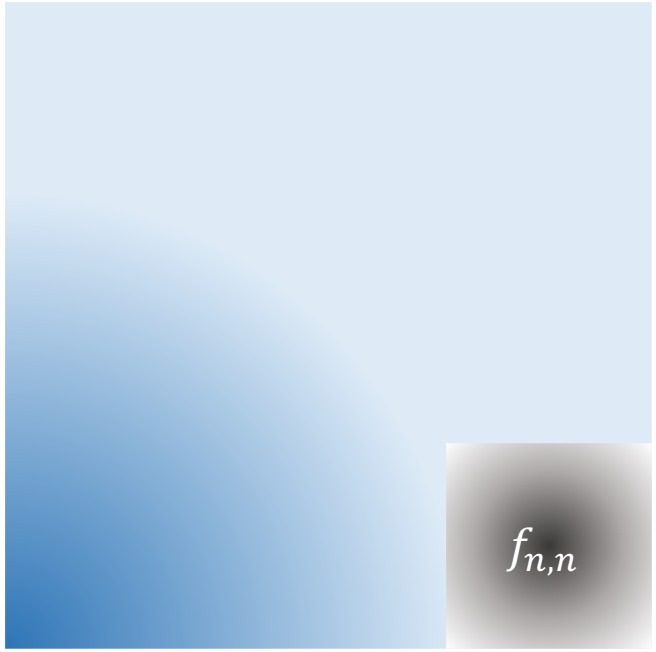
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Spatially-adaptive Low-pass filter



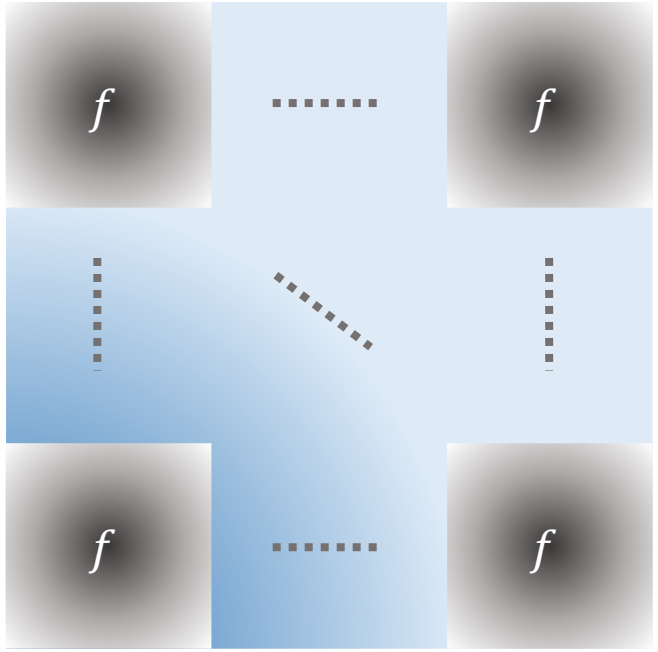
(a) Fixed Lowpass Filter [1]



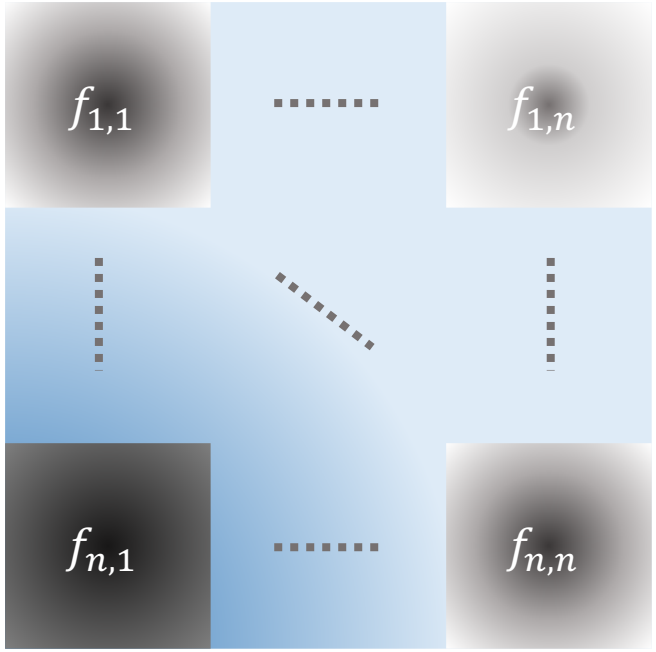
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Low-pass Filter vs Spatial Adaptive Low-pass filter



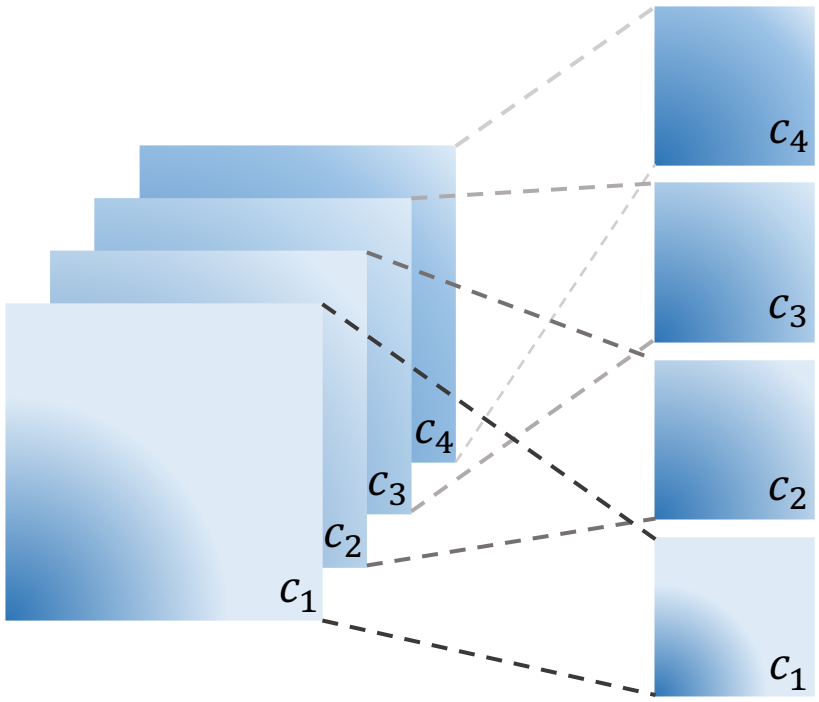
(a) Fixed Lowpass Filter [1]



(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

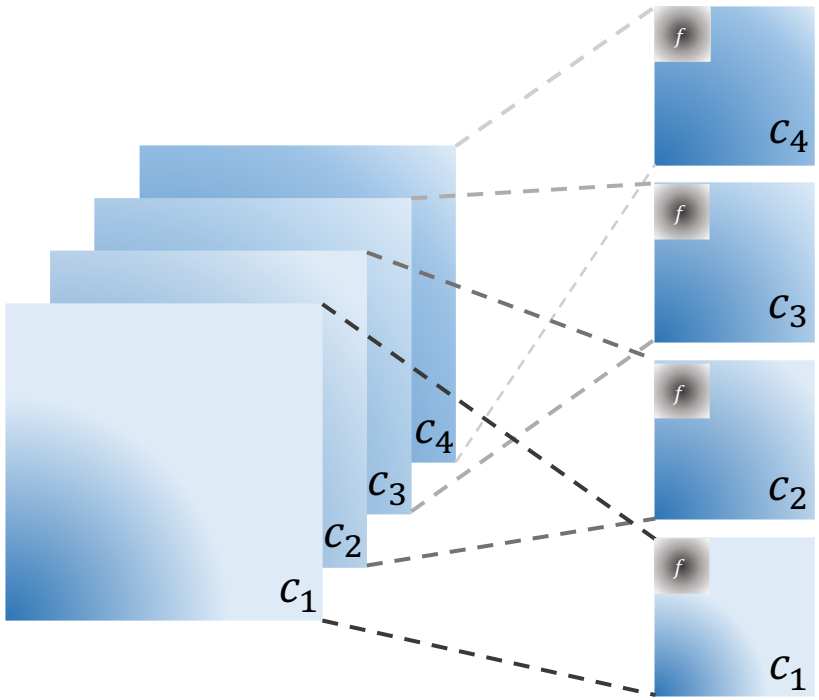
Fixed vs Channel Adaptive Low-pass filter



(a) Fixed Lowpass Filter [1]

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

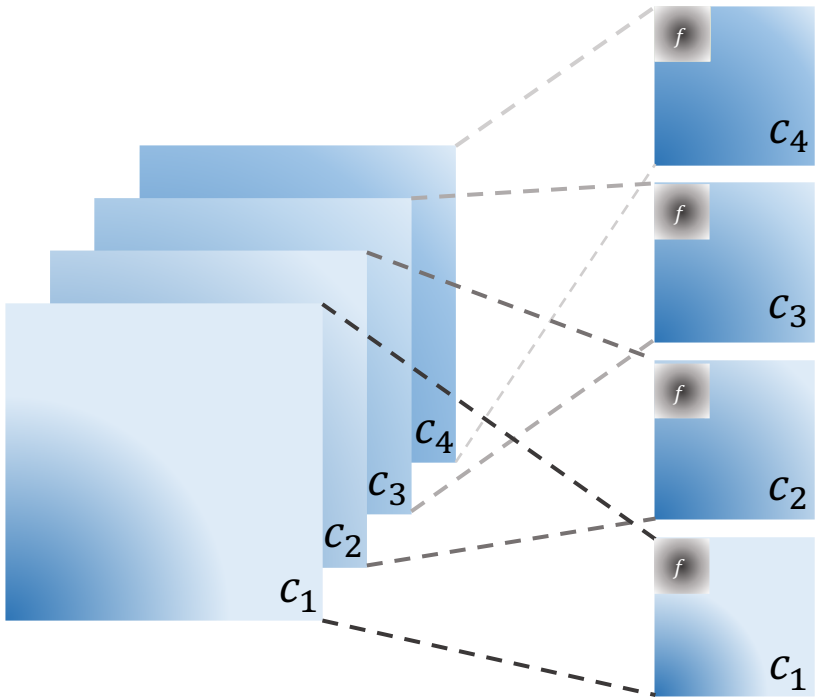
Fixed vs Channel Adaptive Low-pass filter



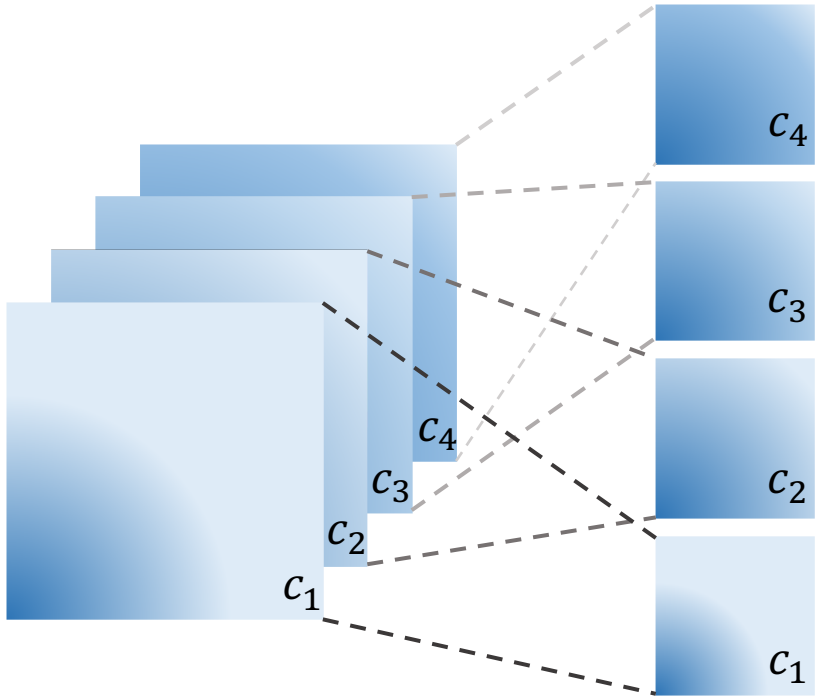
(a) Fixed Lowpass Filter [1]

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Channel Adaptive Low-pass filter



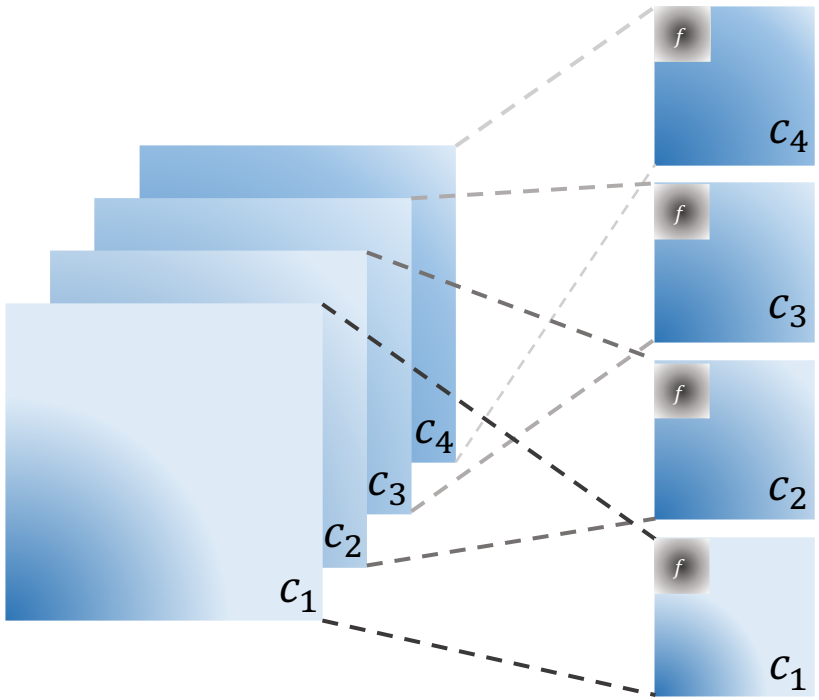
(a) Fixed Lowpass Filter [1]



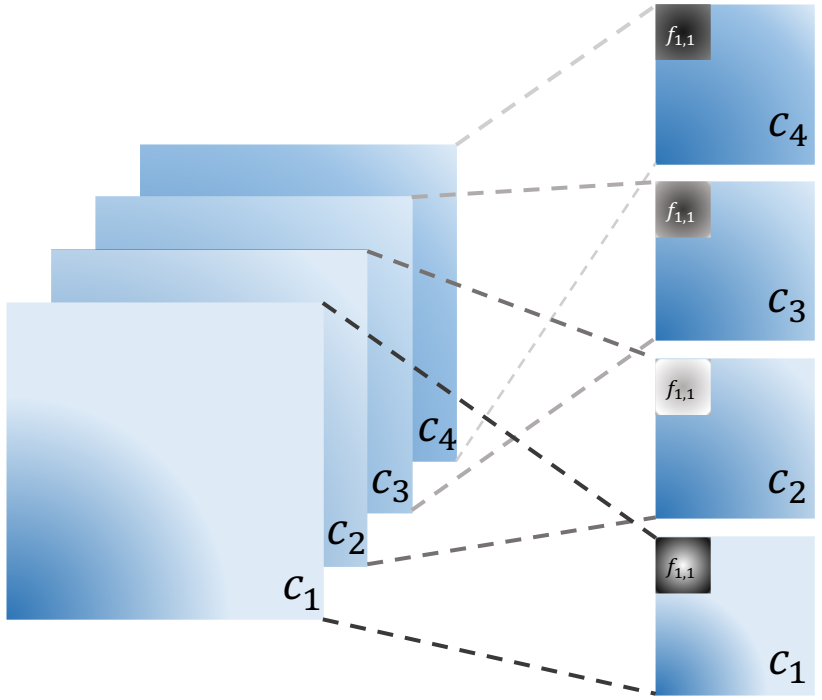
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Channel Adaptive Low-pass filter



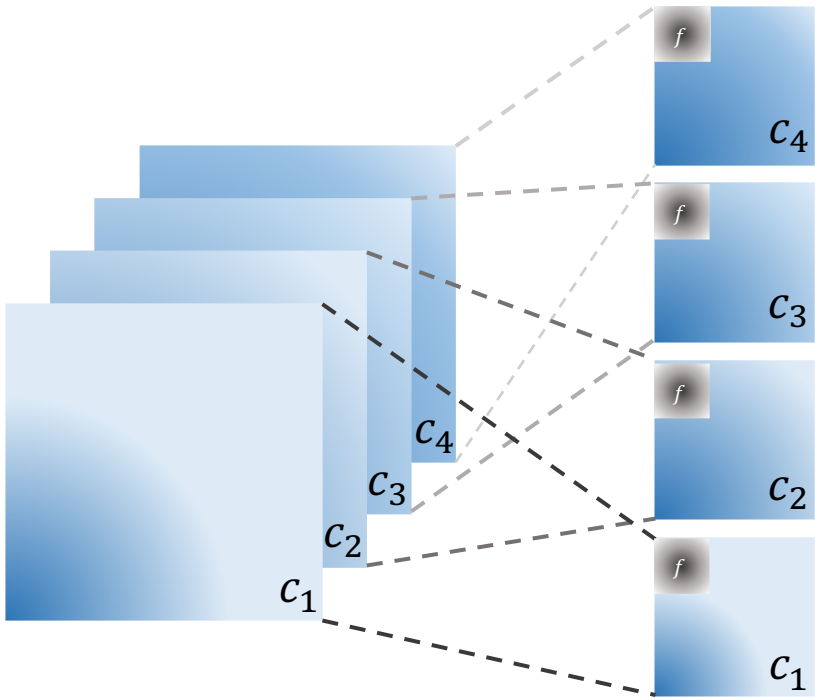
(a) Fixed Lowpass Filter [1]



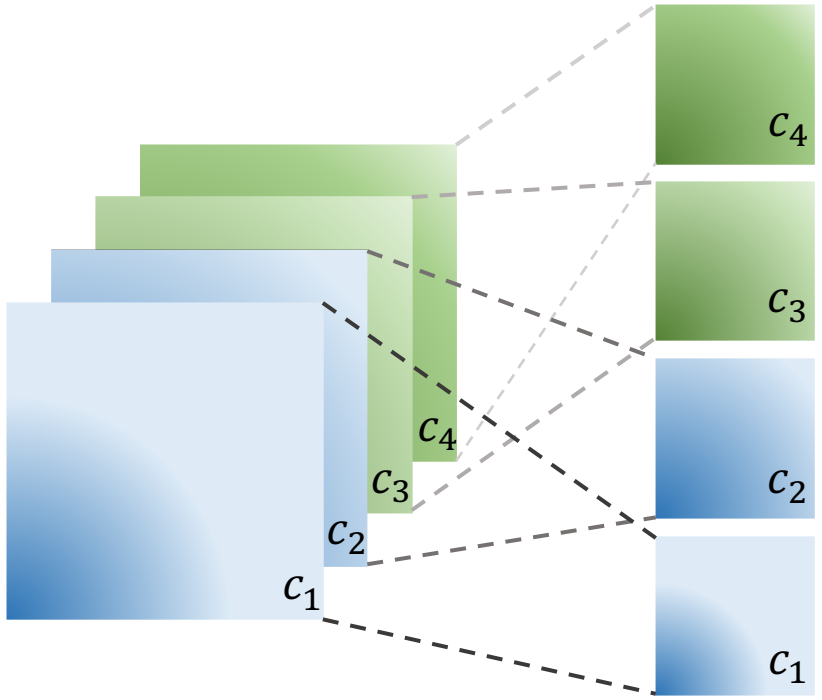
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Group Adaptive Low-pass filter



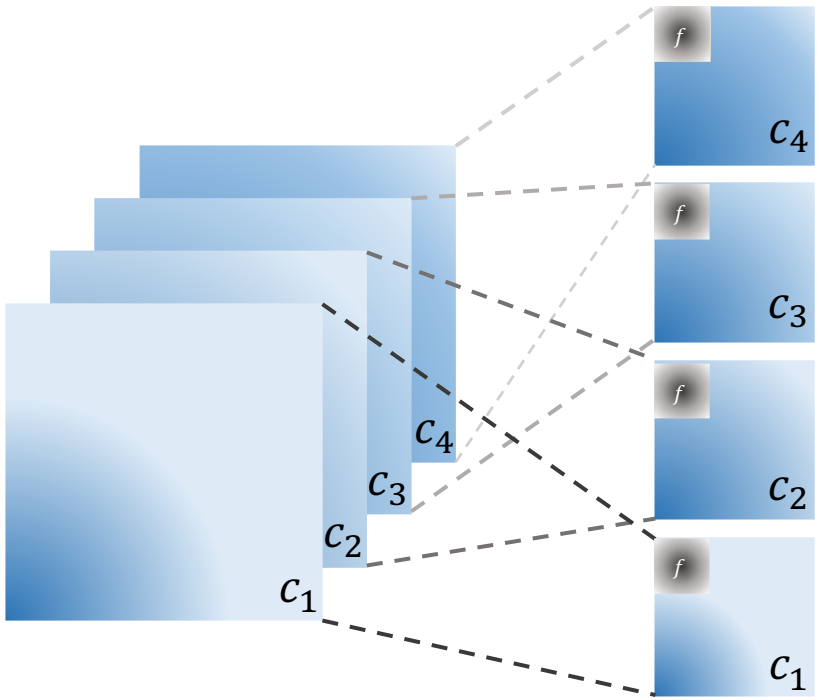
(a) Fixed Lowpass Filter [1]



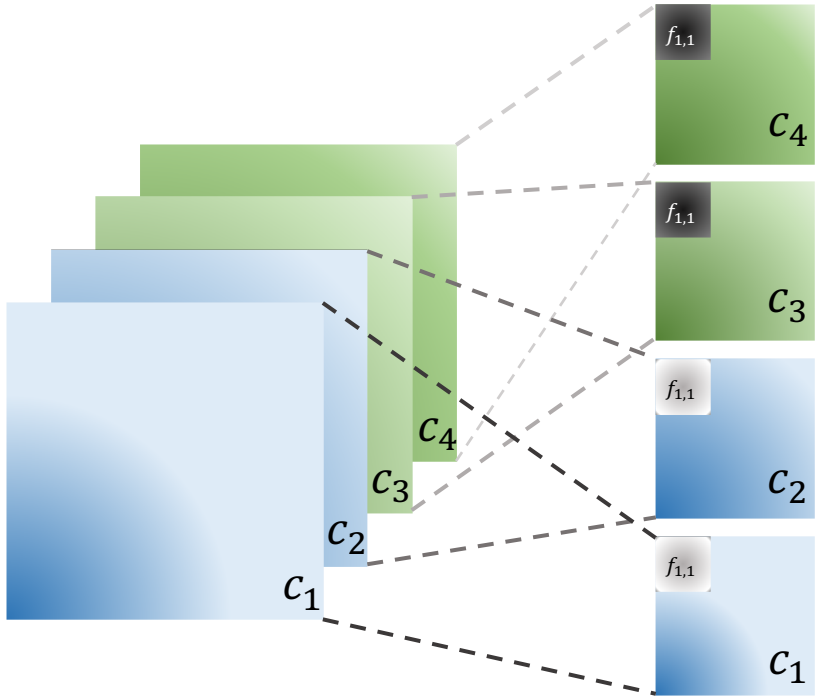
(b) Adaptive Lowpass Filter

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Fixed vs Group Adaptive Low-pass filter



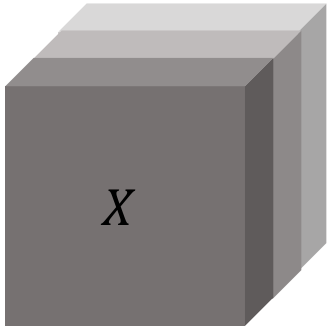
(a) Fixed Lowpass Filter [1]



(b) Adaptive Lowpass Filter

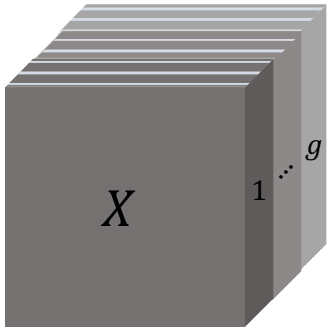
[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

Method Overview

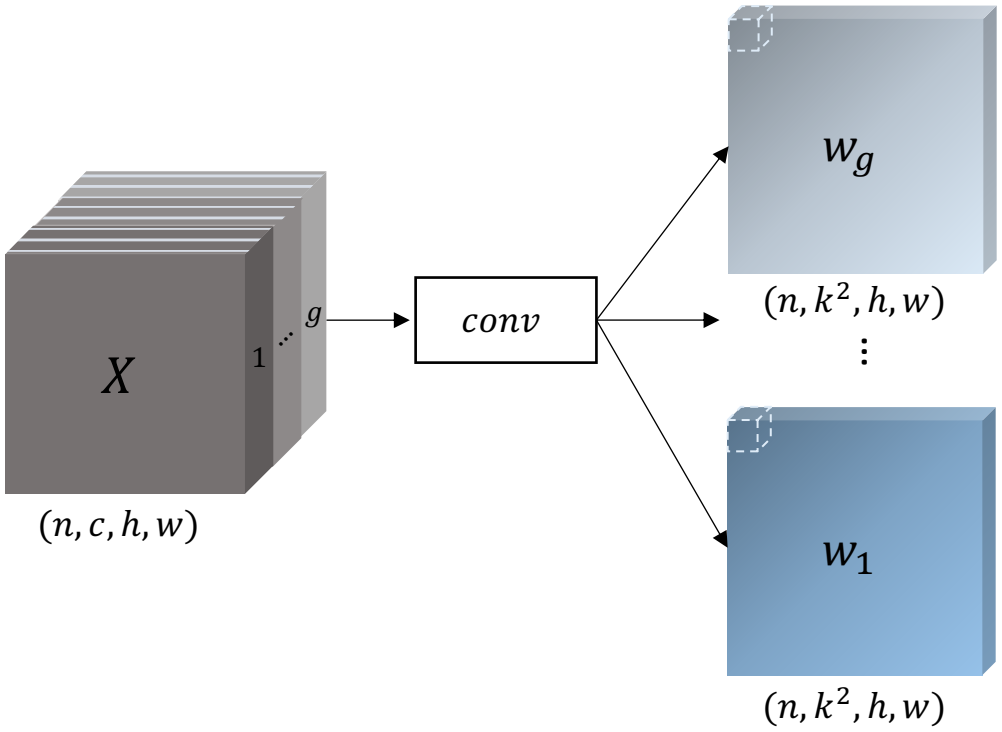


(n, c, h, w)

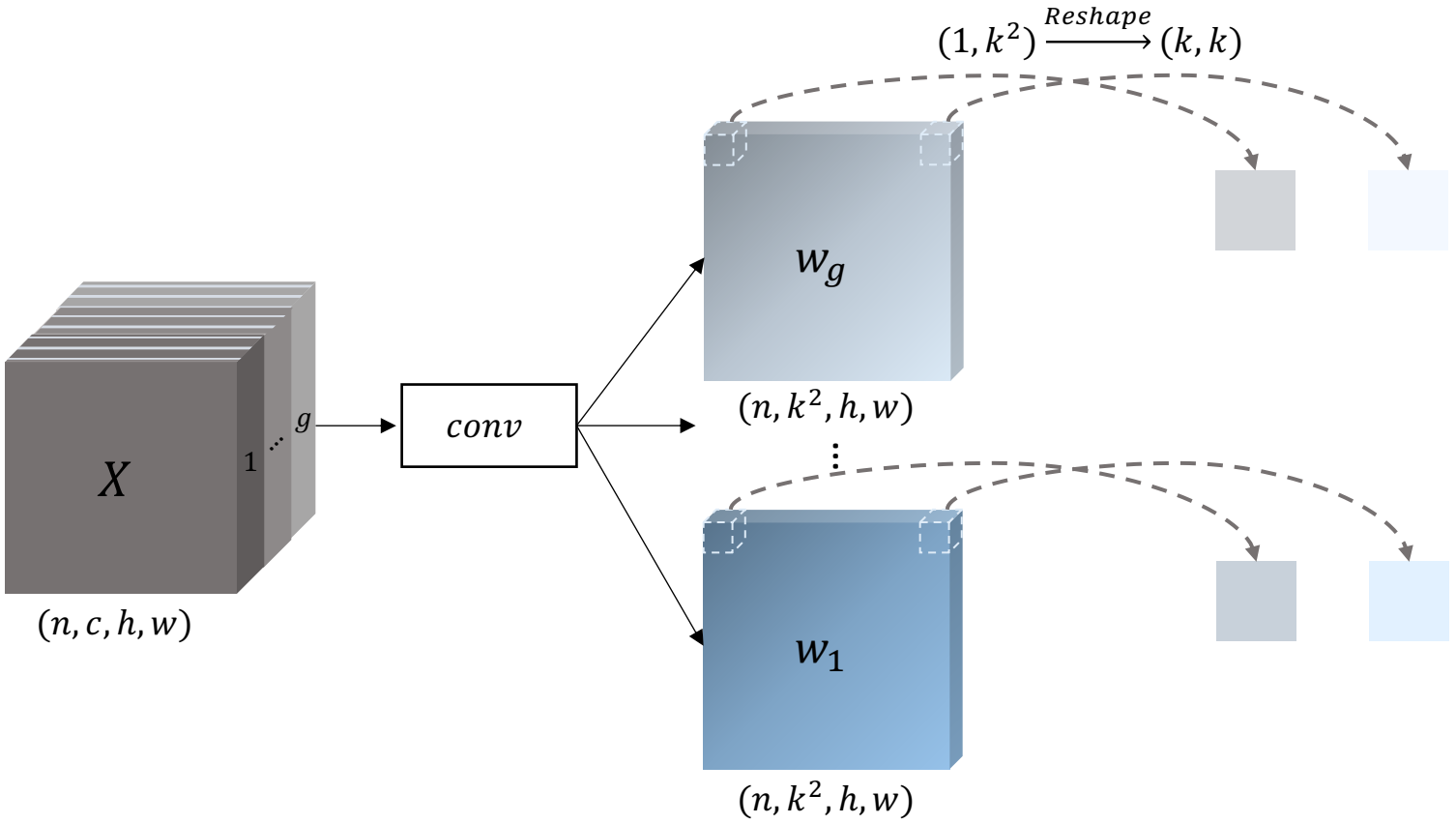
Method Overview



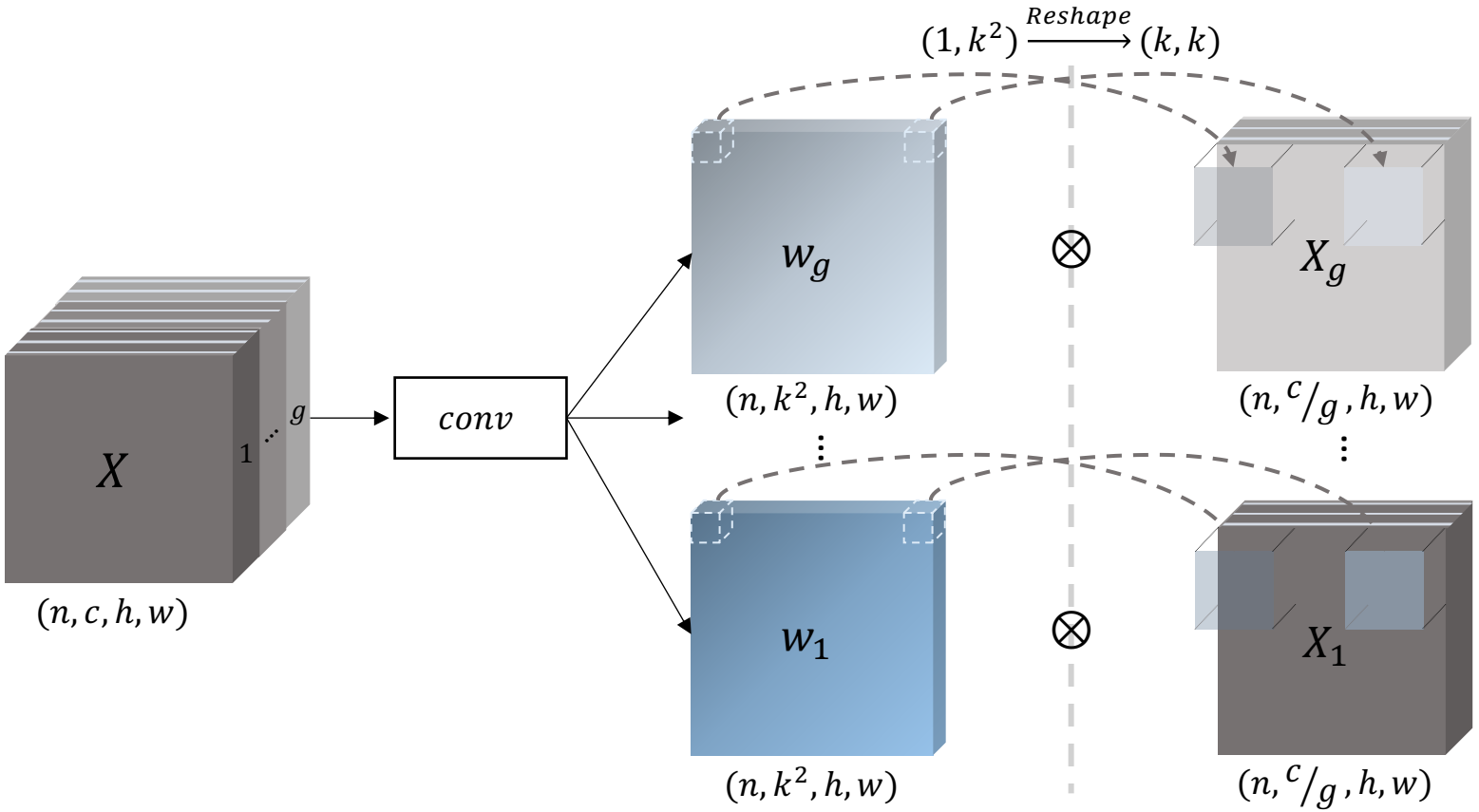
Method Overview



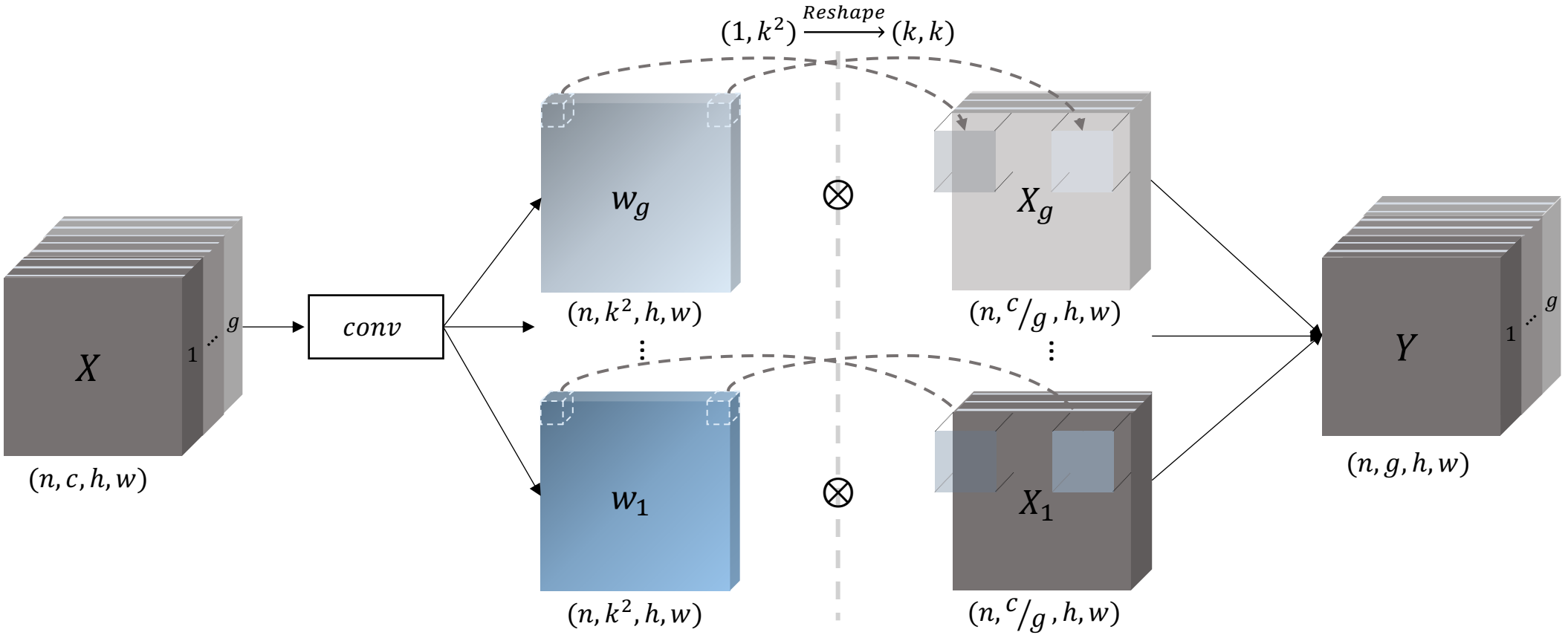
Method Overview



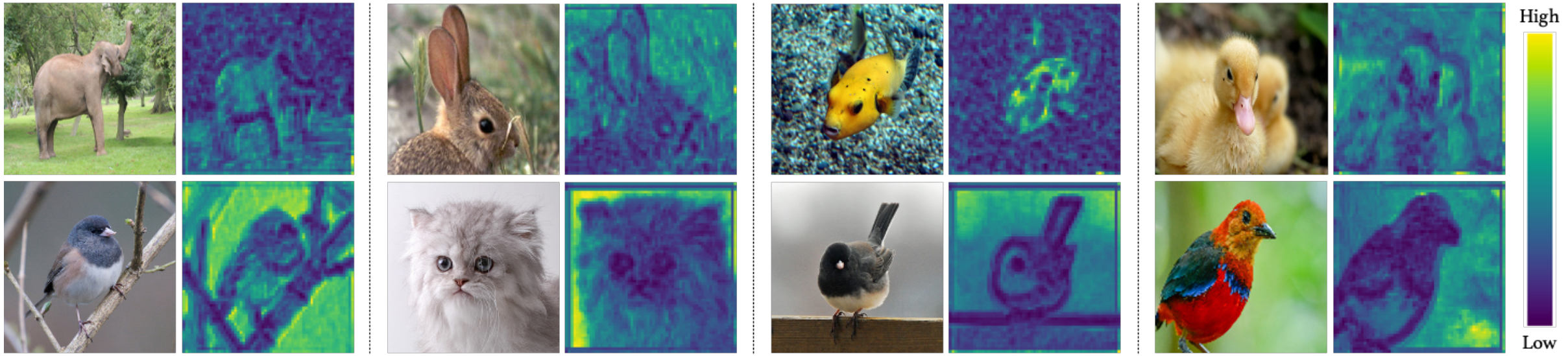
Method Overview



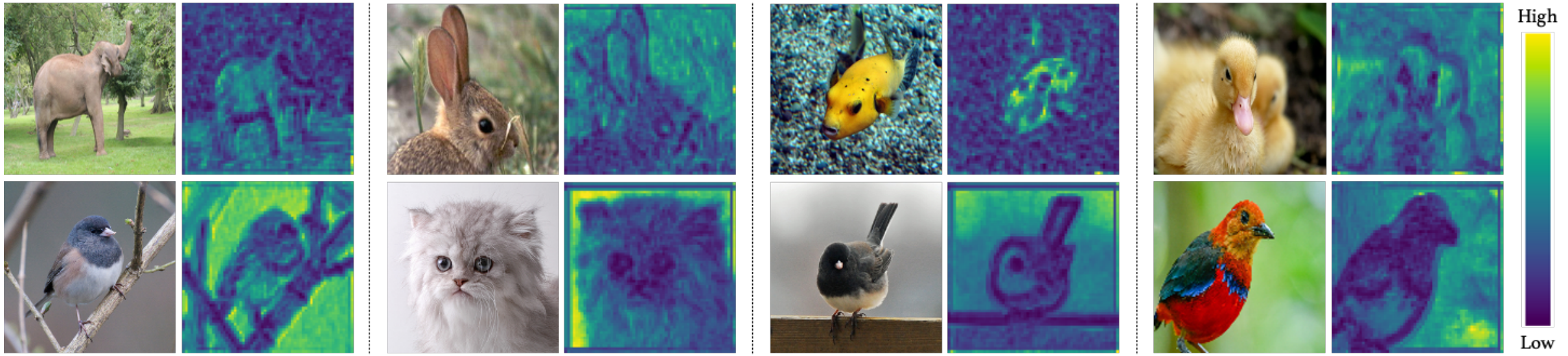
Method Overview



Learned Spatial Filter

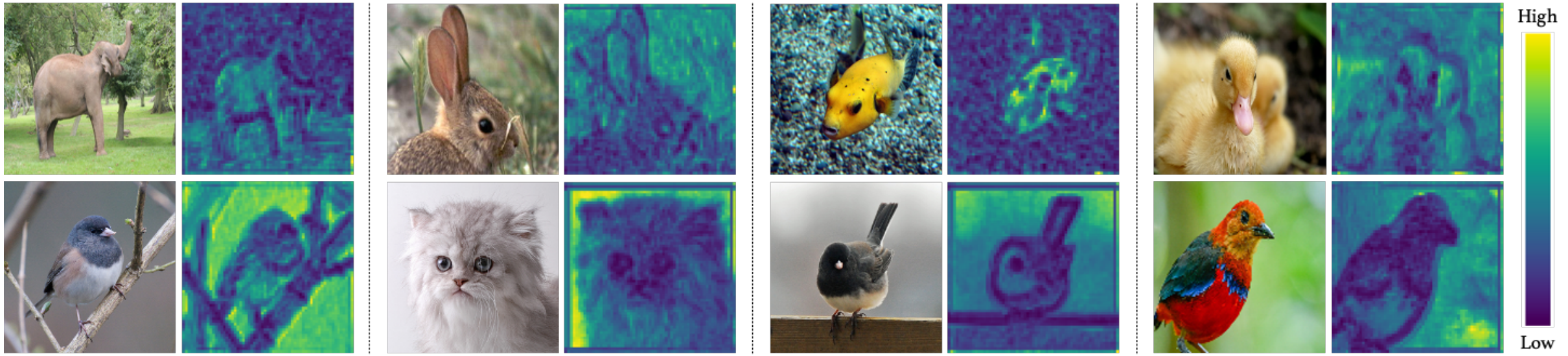


Learned Spatial Filter



Average Filter: $Var($ $\begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array}$ $/\ 9) = 0$ <- Low Variance

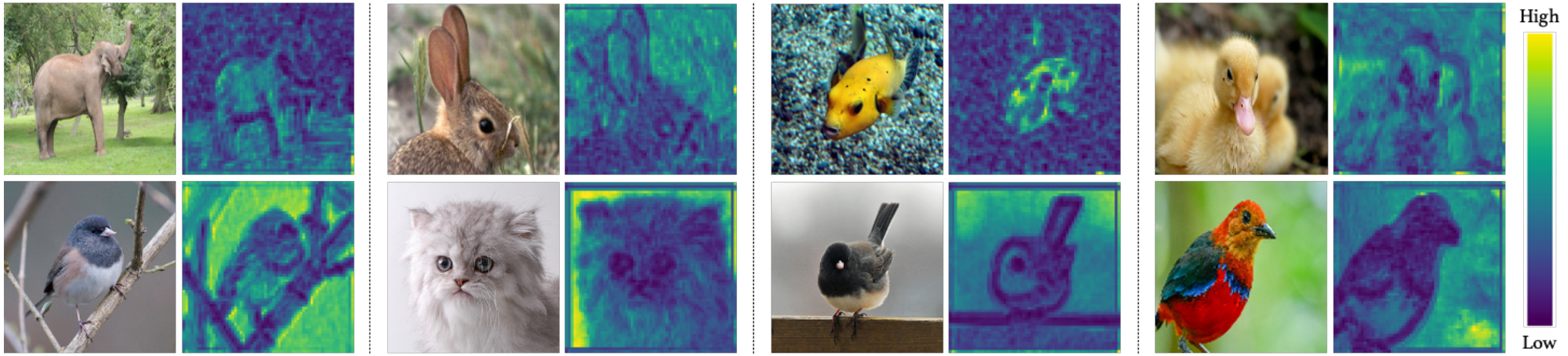
Learned Spatial Filter



Average Filter: $Var\left(\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}/9\right) = 0$ <- Low Variance

Identity Filter: $Var\left(\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}\right) = 0.11$ <- High Variance

Learned Spatial Filter

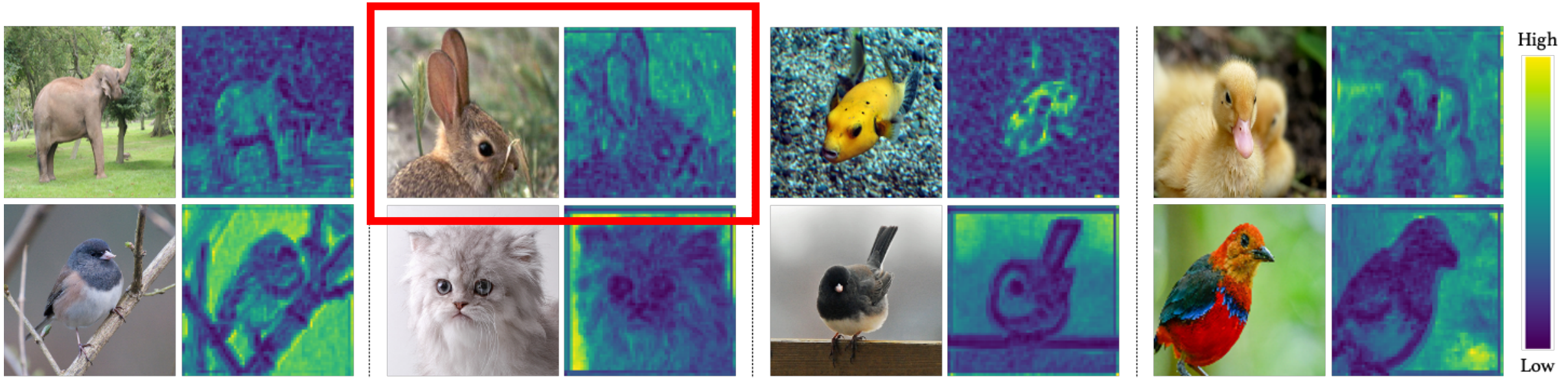


Average Filter: $Var\left(\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}/9\right) = 0$ <- Low Variance

Identity Filter: $Var\left(\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}\right) = 0.11$ <- High Variance

Our model correctly learns to blur high frequency content more to prevent aliasing, and blur low frequency content less to preserve useful information

Learned Spatial Filter

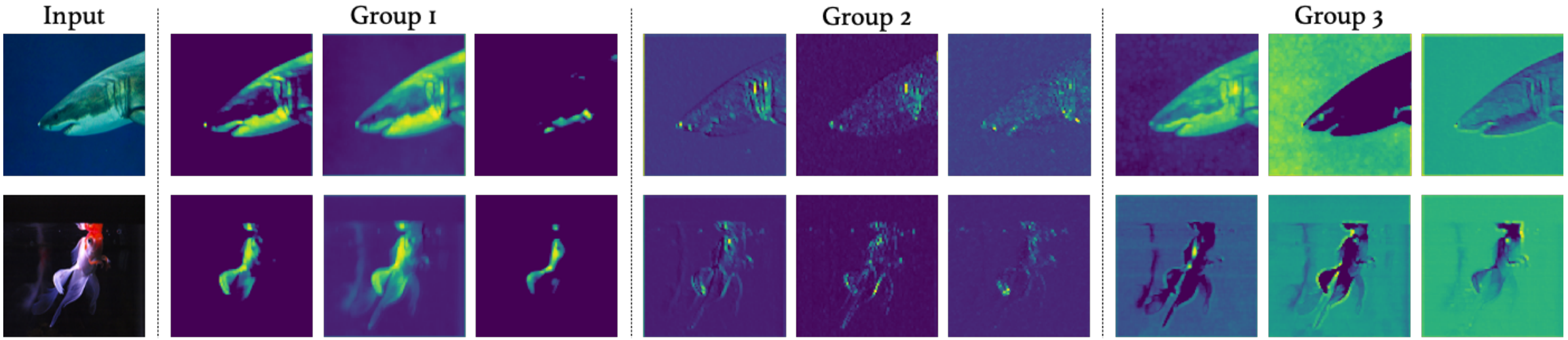


Average Filter: $Var\left(\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}/9\right) = 0$ <- Low Variance

Identity Filter: $Var\left(\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}\right) = 0.11$ <- High Variance

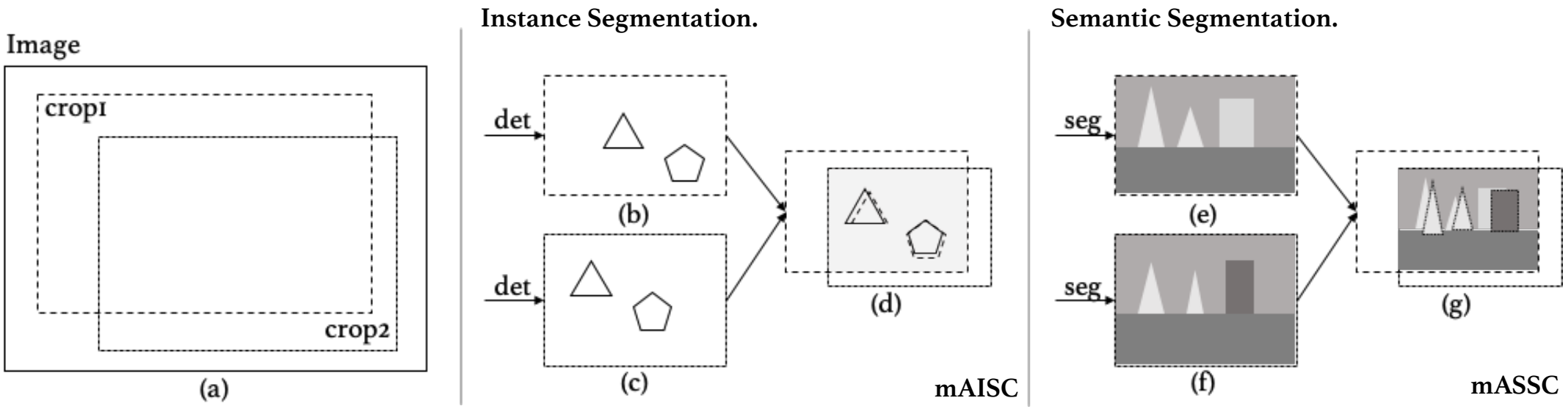
Our model correctly learns to blur high frequency content more to prevent aliasing, and blur low frequency content less to preserve useful information

Learned Group Feature



Features within each group are more similar to each other than those in other groups

Consistency Metric

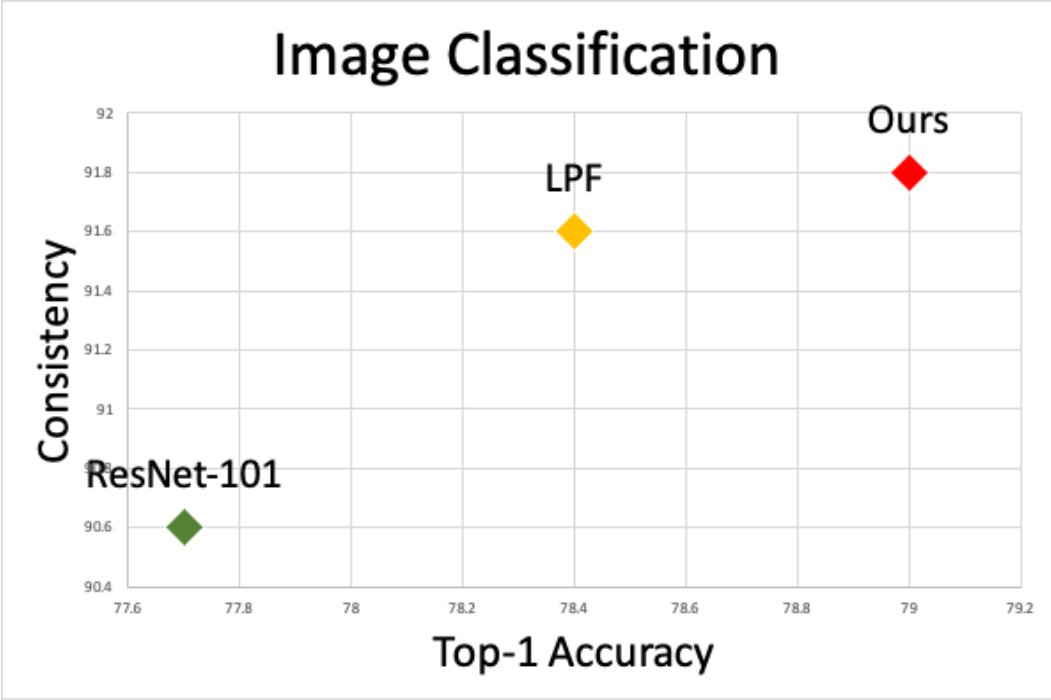


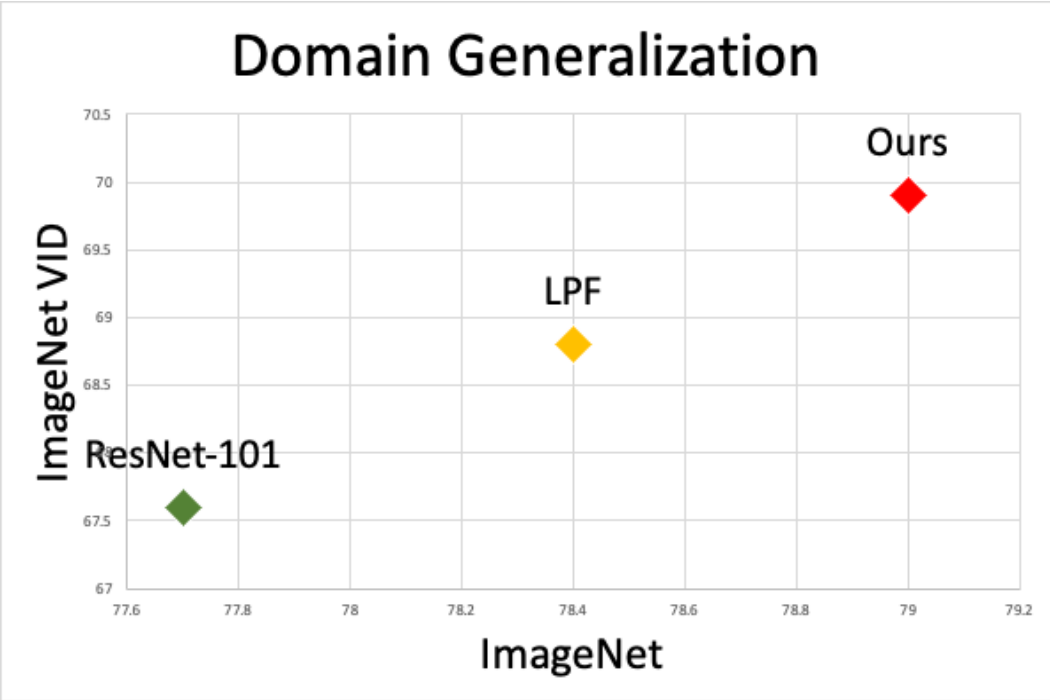
We measure the prediction similarity on the overlapping region of two shifted inputs

Experiment Results

- Image Classification: ImageNet
- Domain Generalization: ImageNet -> ImageNet VID
- Instance Segmentation: COCO
- Semantic Segmentation: PASCAL VOC, Cityscape

Experiment Results

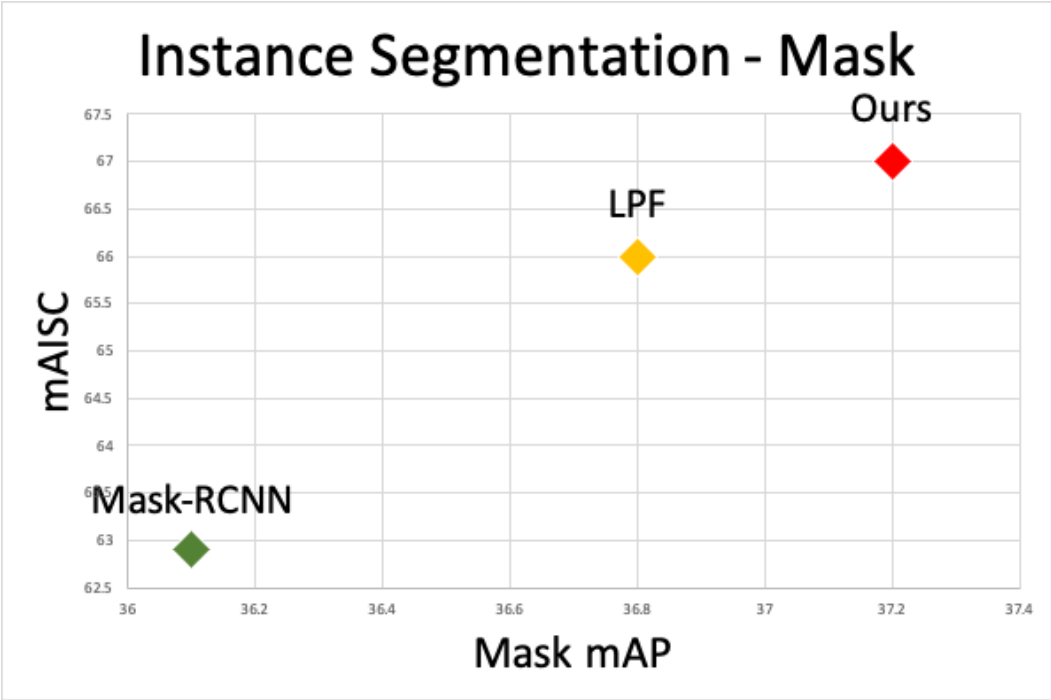


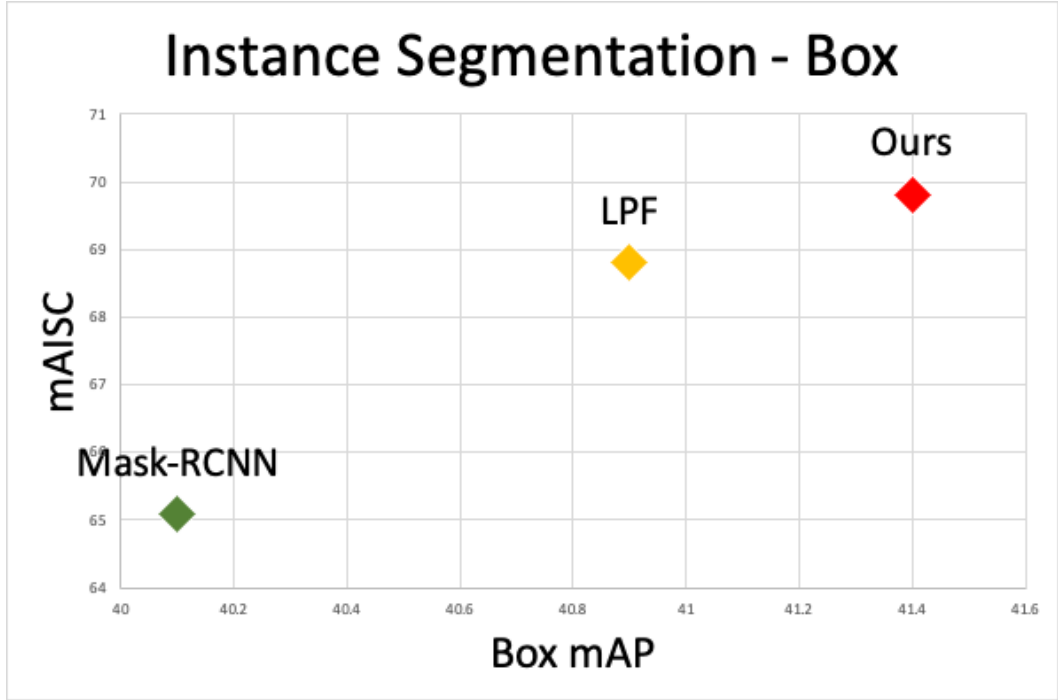


Our method outperforms RensetNet-101 and LPF [1] on both accuracy and consistency

[1] Making Convolution Network Shift and Invariant Again. Zhang. *ICML 2019*

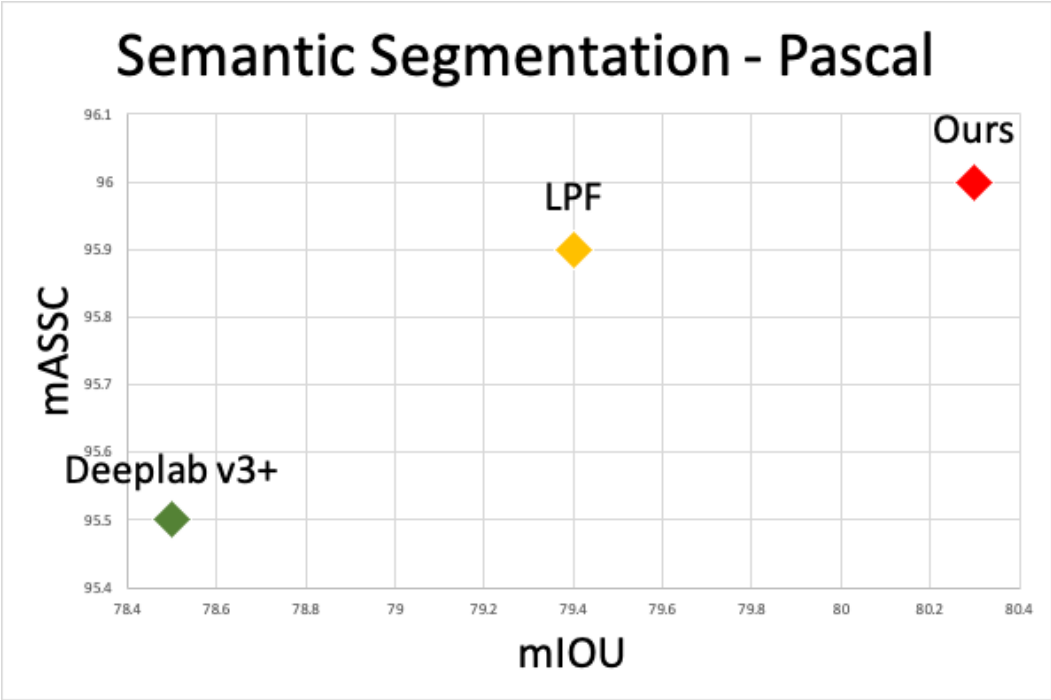
Experiment Results

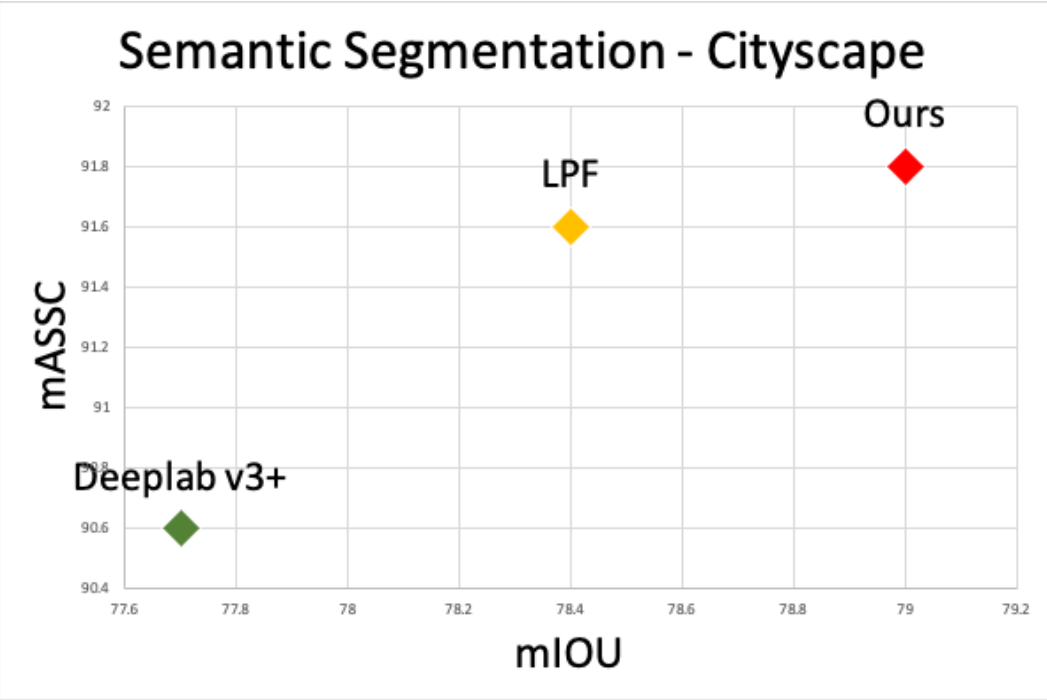




All anti-aliasing techniques improve instance segmentation consistency and accuracy with a large margin

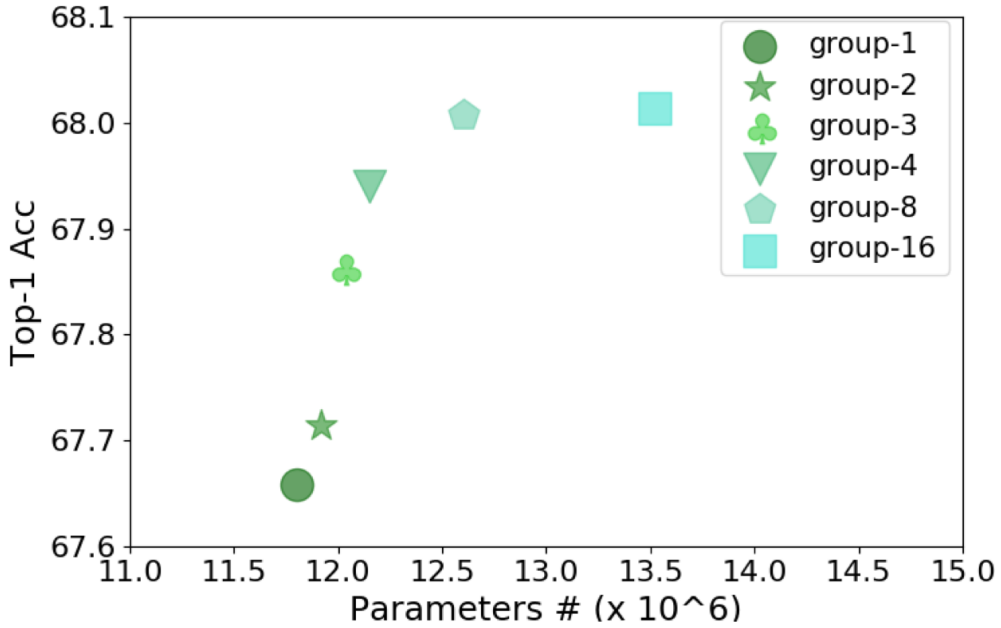
Experiment Results

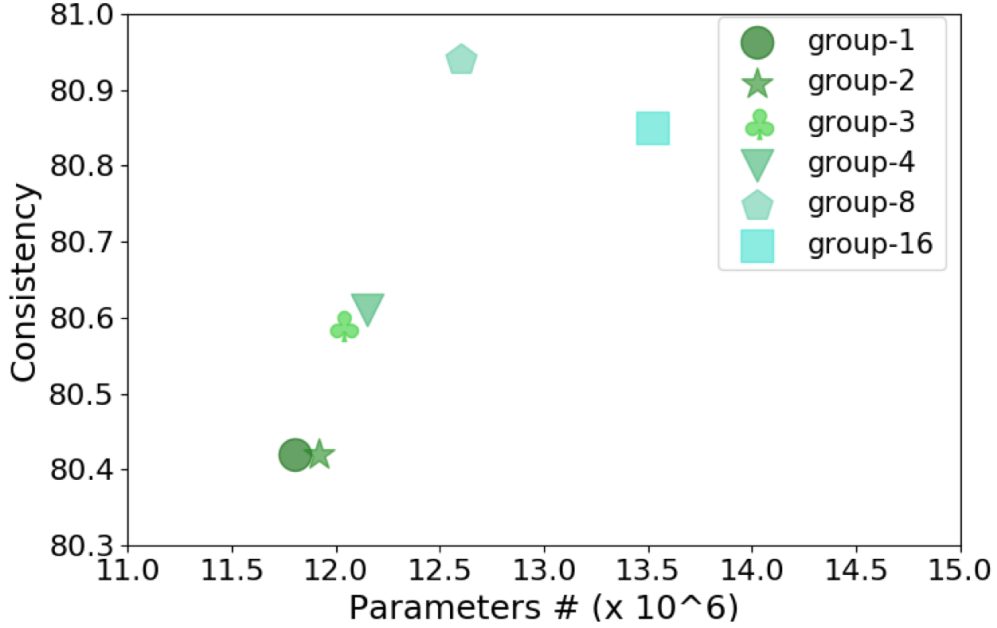




Our method surpass baseline using mIOU metric with 2 points on PASCAL VOC and 1 point on Cityscape

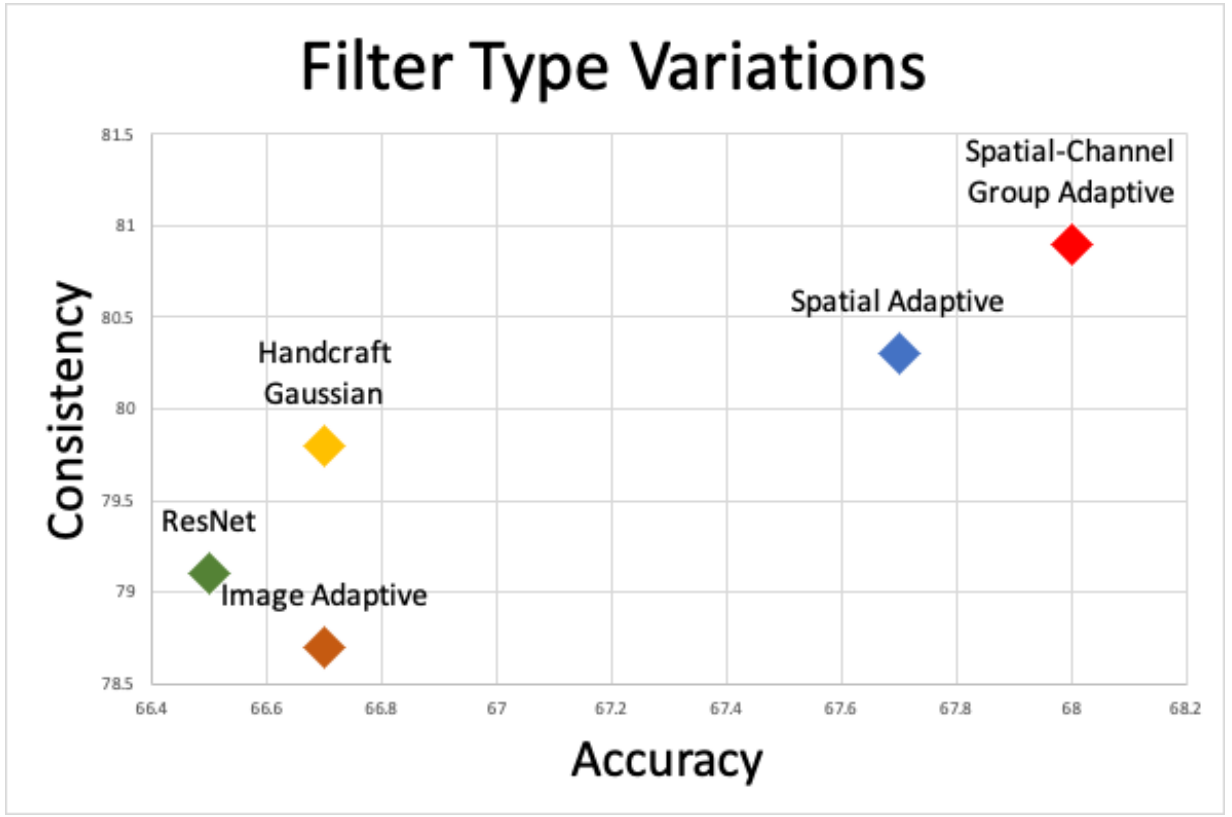
Ablation Study: Group Number





Top-1 Accuracy saturate with group = 8, where consistency drops when group number > 8

Ablation Study: Filter Types



Adaptive spatial and channel filtering method will increase both consistency and accuracy with a large margin

Contributions

- A new adaptive low-pass filtering layer
- New consistency evaluation metrics on pixel classification tasks
- Evaluate our approach on ResNet-101, Mask-RCNN, Deeplab v3+

Thank You!