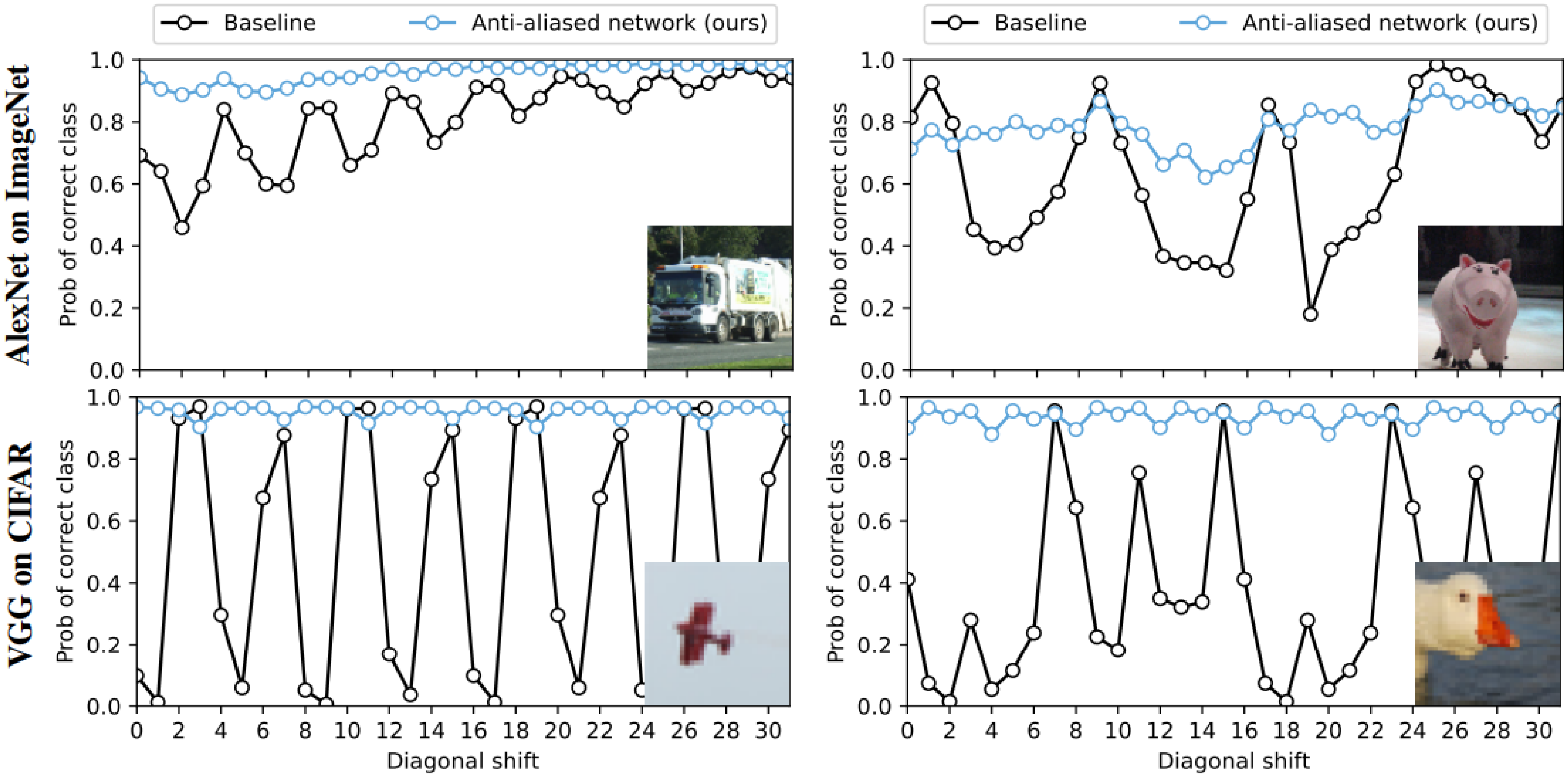
Making Convolutional Networks Shift-Invariant Again

Richard Zhang [1]

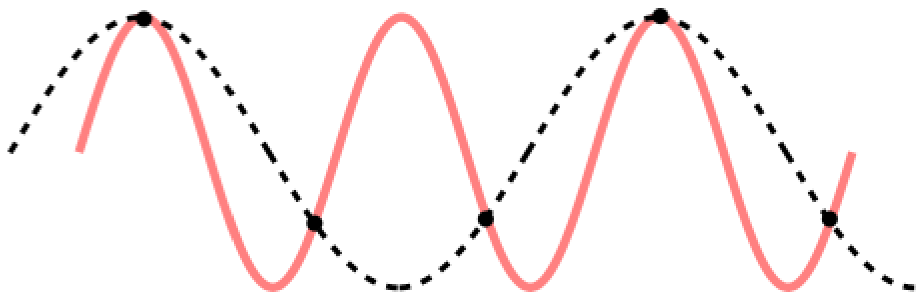
ICML, 2019

Presented by : Kangyeol Kim
DAVIAN Lab, Korea University

Is CNN shift-invariant?

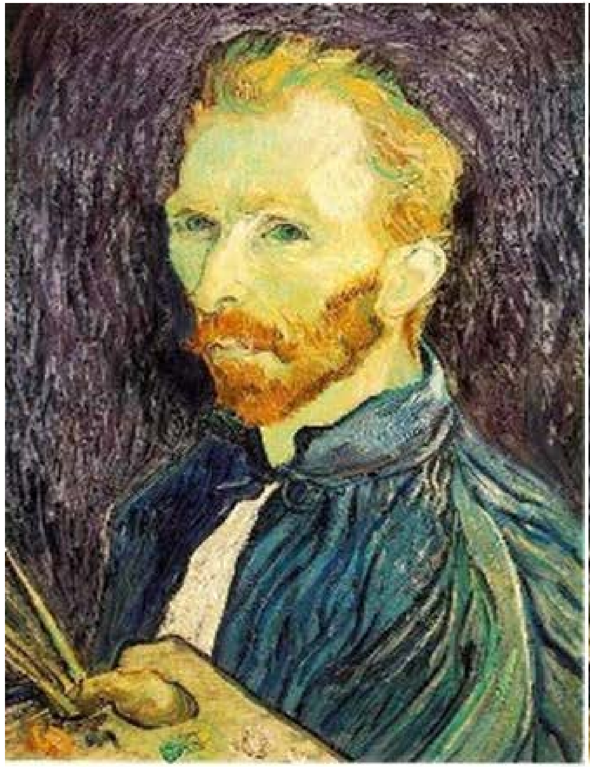


- **Naïve Maxpooling effects on image**

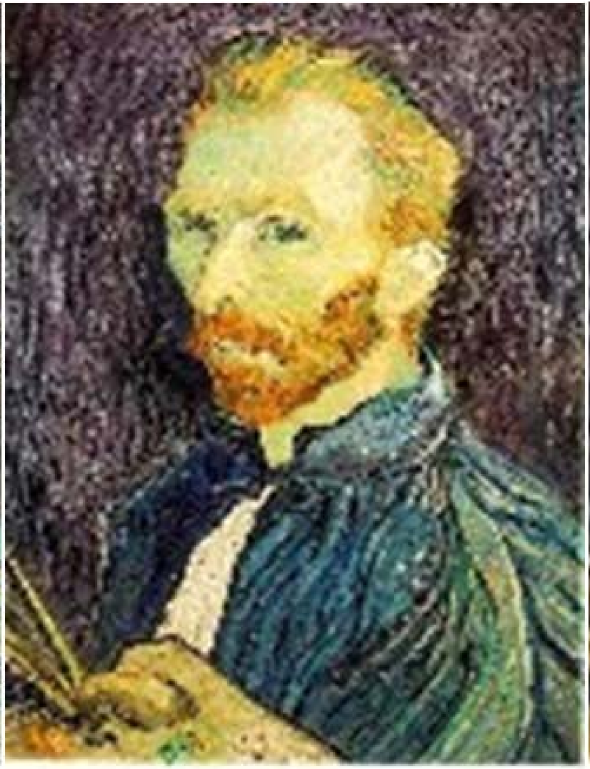


Problem of subsampling in CNN

- **Naïve Maxpooling effects on image**



1/2

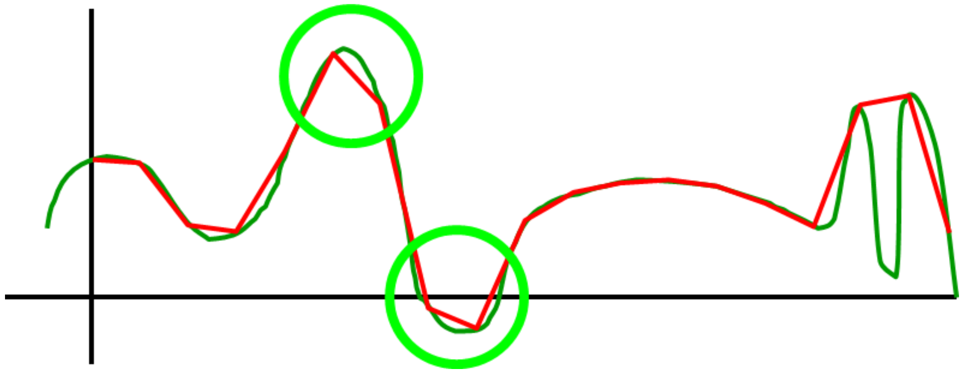


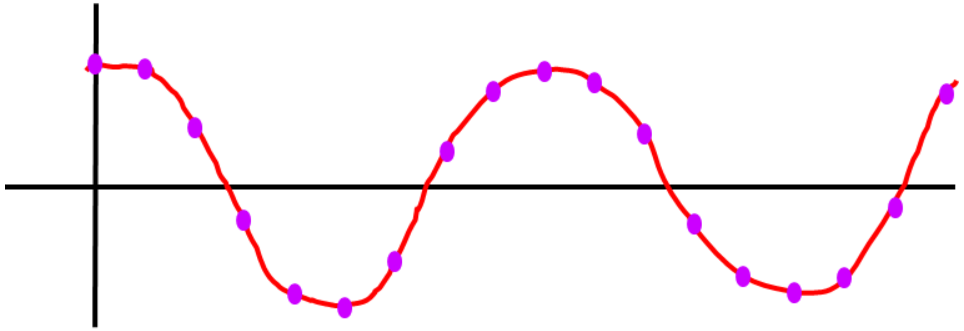
1/4 (2x zoom)

1/8 (4x zoom)

Conventional solution

- **Smoothing**

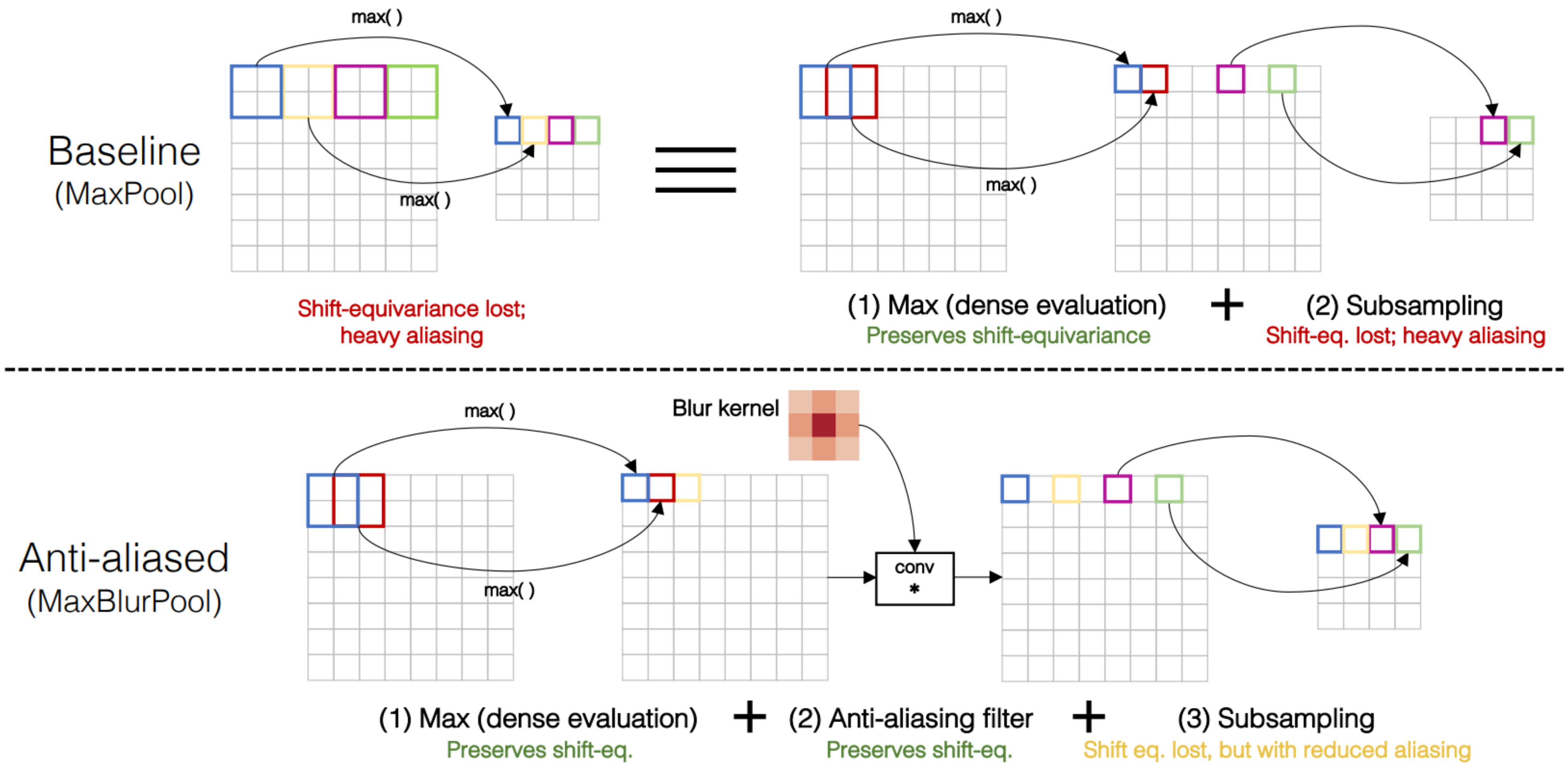




Conventional solution

- **Smoothing, similar to average pooling**

Proposed methods, Blur function



Proposed methods, 1D example

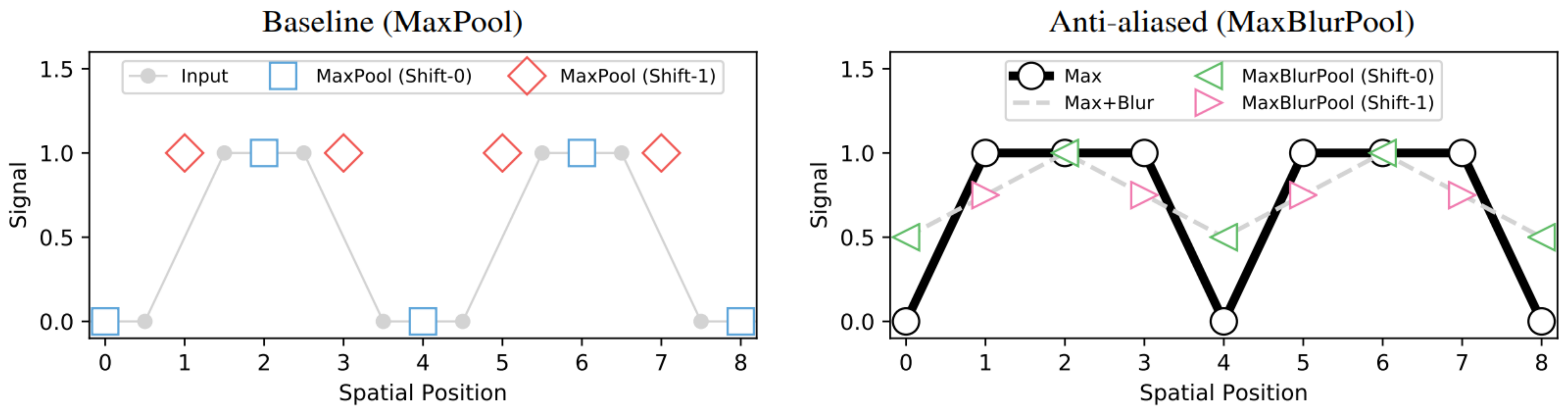


Figure 4. **Illustrative 1-D example of sensitivity to shifts.** We illustrate how downsampling affects shift-equivariance with a toy example. **(Left)** An input signal is in light gray line. Max-pooled ($k = 2$, $s = 2$) signal is in blue squares. Simply shifting the input and then max-pooling provides a completely different answer (red diamonds). **(Right)** The blue and red points are subsampled from a densely max-pooled ($k = 2$, $s = 1$) intermediate signal (**thick black line**). We low-pass filter this intermediate signal and then subsample from it, shown with green and magenta triangles, better preserving shift-equivariance.

Proposed methods, Choice of filter

- ***Rectangle-2*** [1, 1]: moving average or box filter; equivalent to average pooling or "nearest" downsampling
- ***Triangle-3*** [1, 2, 1]: two box filters convolved together; equivalent to bilinear downsampling
- ***Binomial-5*** [1, 4, 6, 4, 1]: the box filter convolved with itself repeatedly; the standard filter used in Laplacian pyramids (Burt & Adelson, 1987)

Proposed methods, Conv/AvgPool

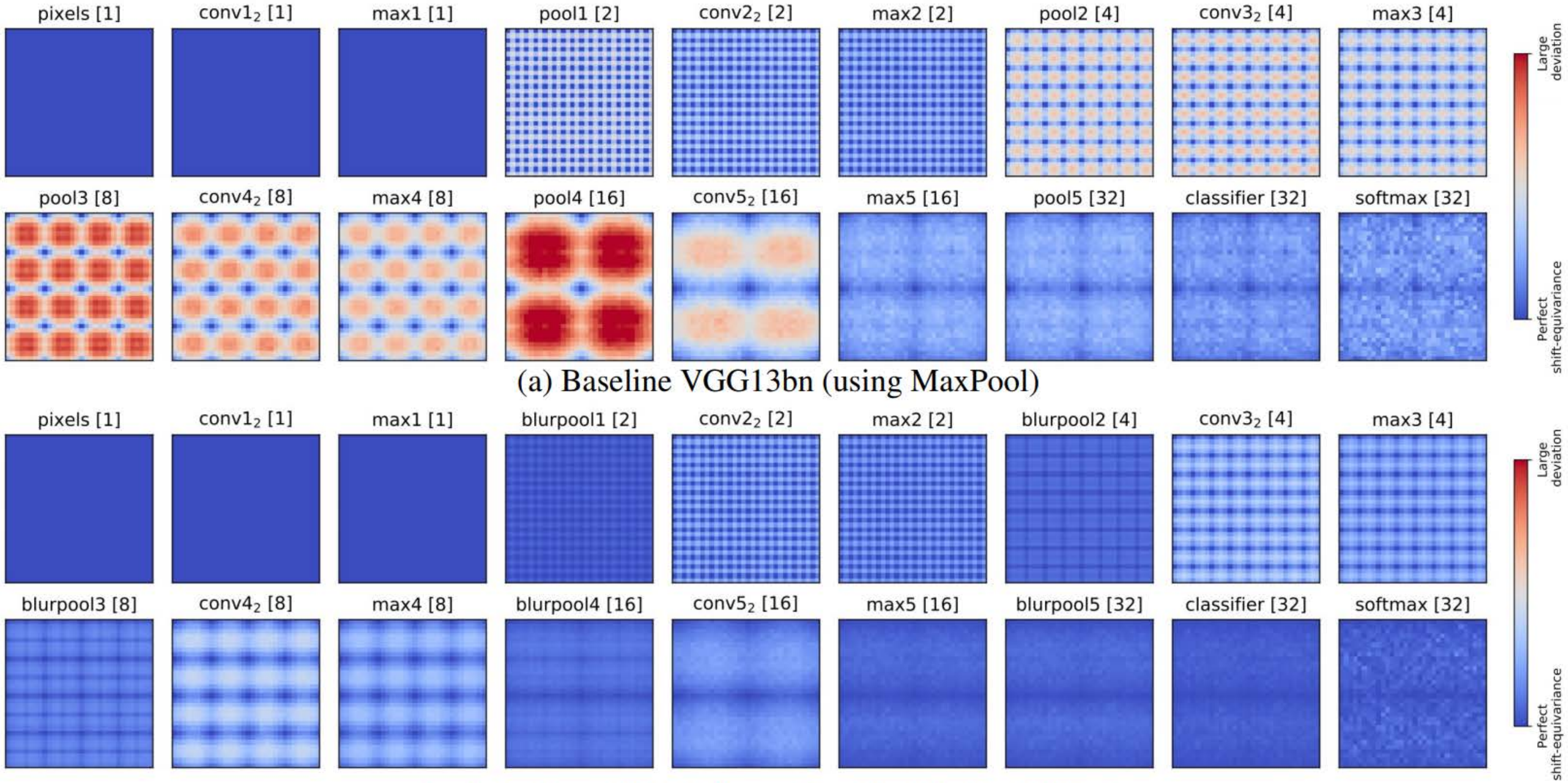
StridedConv→ConvBlurPool Strided-convolutions suffer from the same issue, and the same method applies.

$$\text{Relu} \circ \text{Conv}_{k,s} \rightarrow \text{BlurPool}_{m,s} \circ \text{Relu} \circ \text{Conv}_{k,1} \quad (5)$$

AveragePool→BlurPool Blurred downsampling with a box filter is the same as average pooling. Replacing it with a stronger filter provides better shift-equivariance. We examine such filters next.

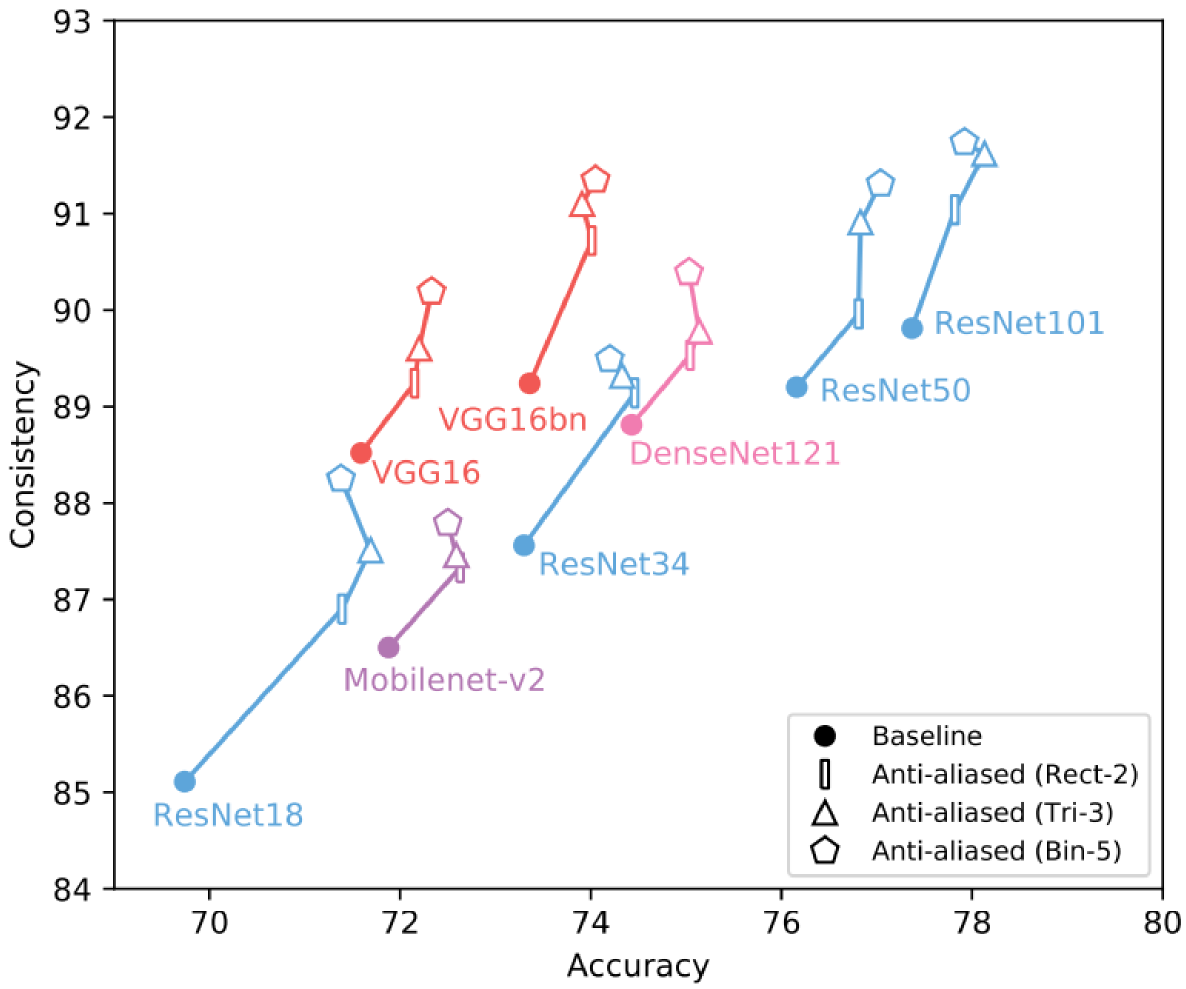
$$\text{AvgPool}_{k,s} \rightarrow \text{BlurPool}_{m,s} \quad (6)$$

Experiment, Feature distance



(b) Anti-aliased VGG13bn (using MaxBlurPool, *Bin-5*)

Experiment, Improved performance



Experiment, Improved performance

	Normalized average		Unnormalized average	
	ImNet-C	ImNet-P	ImNet-C	ImNet-P
	mCE	mFR	mCE	mFR
Baseline	76.4	58.0	60.6	7.92
Rect-2	75.2	56.3	59.5	7.71
Tri-3	73.7	51.9	58.4	7.05
Bin-5	**73.4**	**51.2**	**58.1**	**6.90**

Table 2. **Accuracy and stability robustness.** Accuracy in ImageNet-C, which contains systematically corrupted ImageNet images, measured by mean corruption error **mCE** (lower is better). Stability on ImageNet-P, which contains perturbed image sequences, measured by mean flip rate **mFR** (lower is better). We show raw, unnormalized scores, as well as scores normalized to AlexNet, as used in Hendrycks et al. (2019). Anti-aliasing improves both accuracy and stability over the baseline. All networks are variants of ResNet50.