Shift Invariant Module

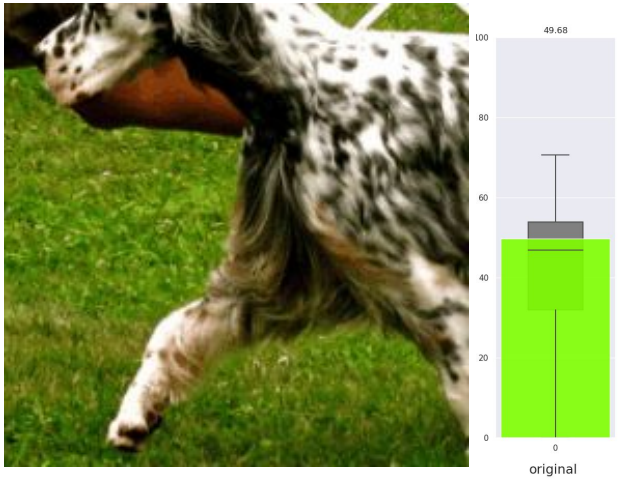
Nian-Hsuan Tsai, Fong-An Chang,
Mu-Chien Hsu, Kai-Hsiang Liu

Motivation

CNNs are shift invariant…right?

No, it's actually not!

But why?



Problem

[1] [1904.11486] Making Convolutional Networks Shift-Invariant Again (arxiv.org)
[2] Making Convolutional Networks Shift-Invariant Again (richzhang.github.io)

Aliasing

The problem lies in downsampling!

Take max pooling as an example:

Original: 00 11 00 11 (shift-0)
→ 0 1 0 1

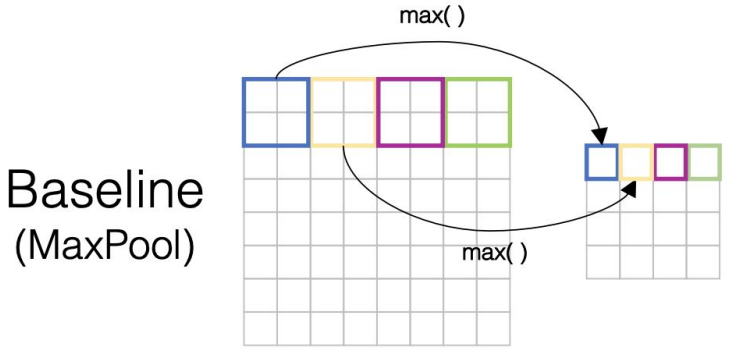
Shifted: 01 10 01 10 (shift-1)
→ 1 1 1 1

Aliasing

When does aliasing happen?

- Max pool
- Average pool
- Strided convolutions

How to fix this?

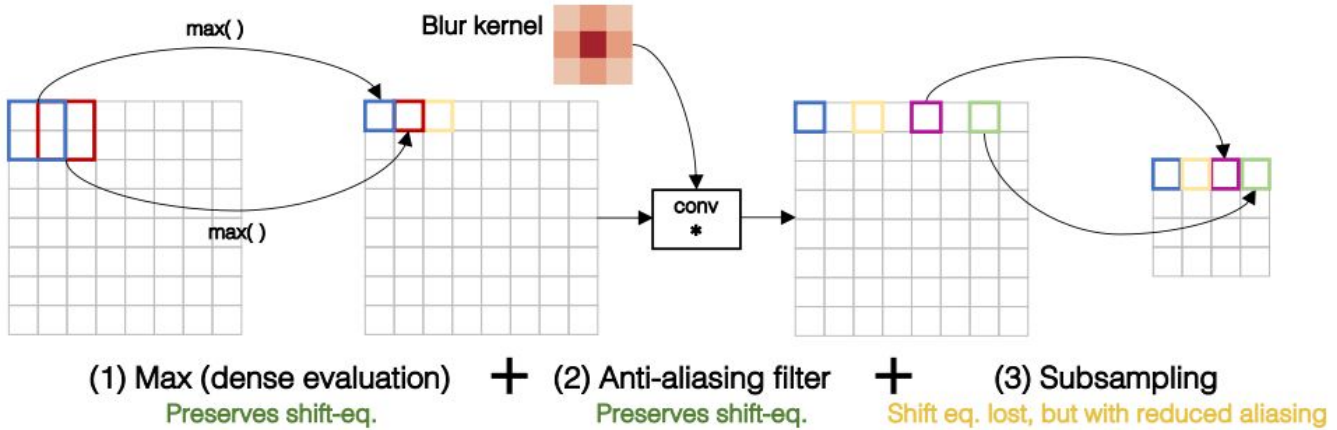


Prior Work

[1] [1904.11486] Making Convolutional Networks Shift-Invariant Again (arxiv.org)

AACNN: Add Blur!

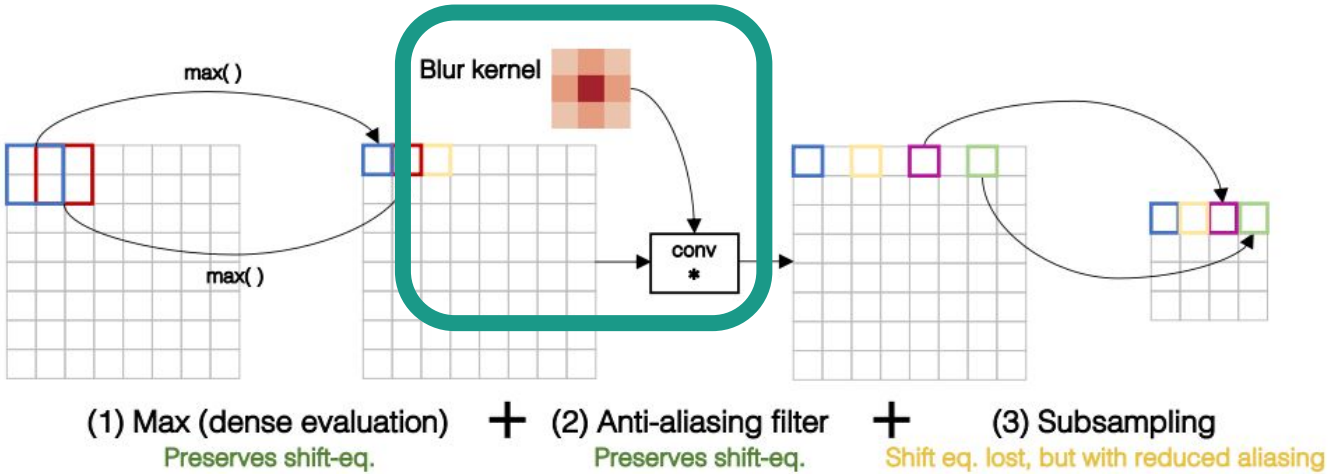
Anti-aliased
(MaxBlurPool)



R. Zhang. **Making Convolutional Networks Shift-Invariant Again.** In ICML, 2019.

Why not learn the aggregation?

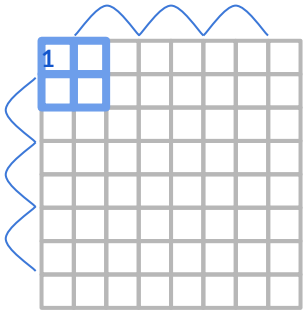
Anti-aliased
(MaxBlurPool)



Proposed Method

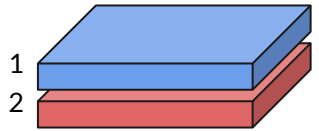
Offsetted Operations

x offset=0, y offset=0, stride=2

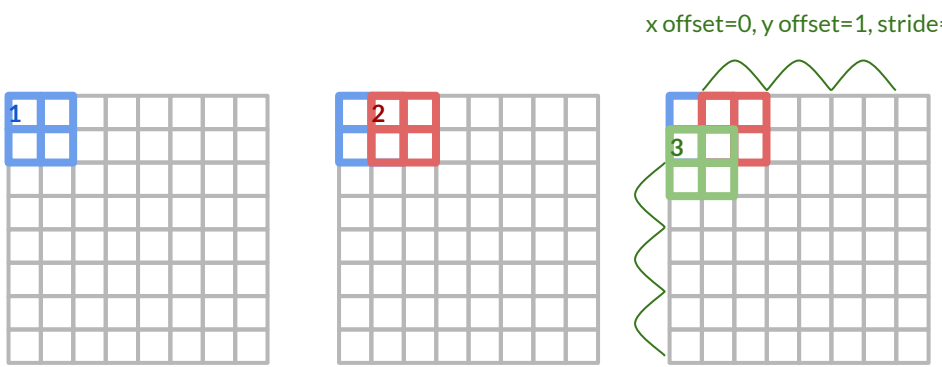


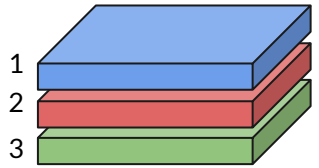
1 * H * W

Offsetted Operations

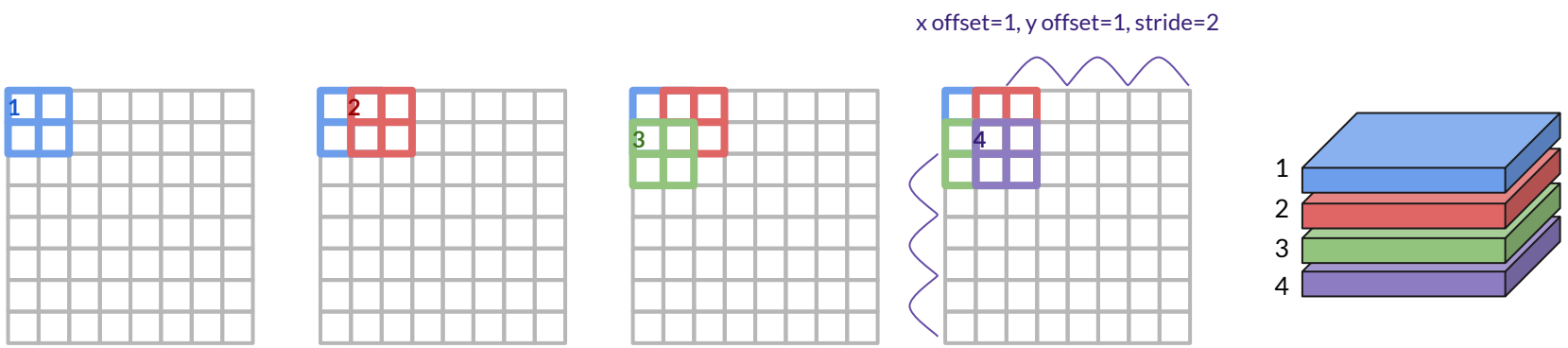


Offsetted Operations





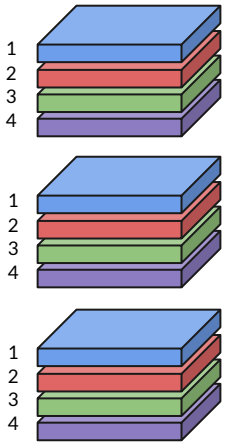
Offsetted Operations



1 * H * W

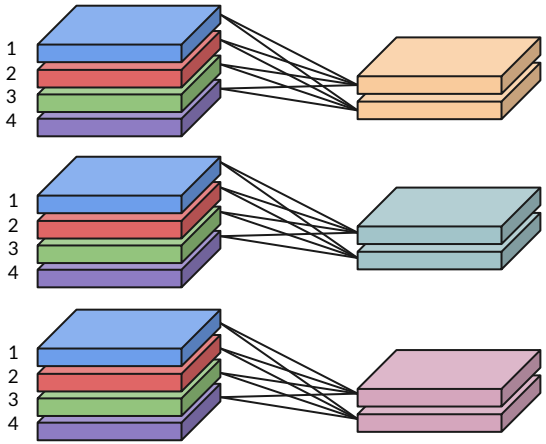
4 * H/2 * W/2

Multiple Channels

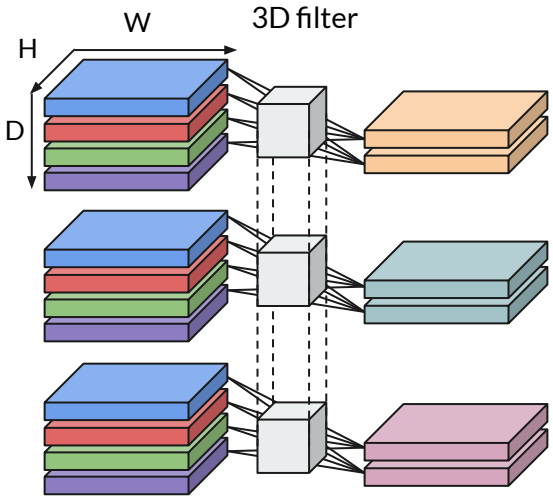


C * 4 * H/2 * W/2

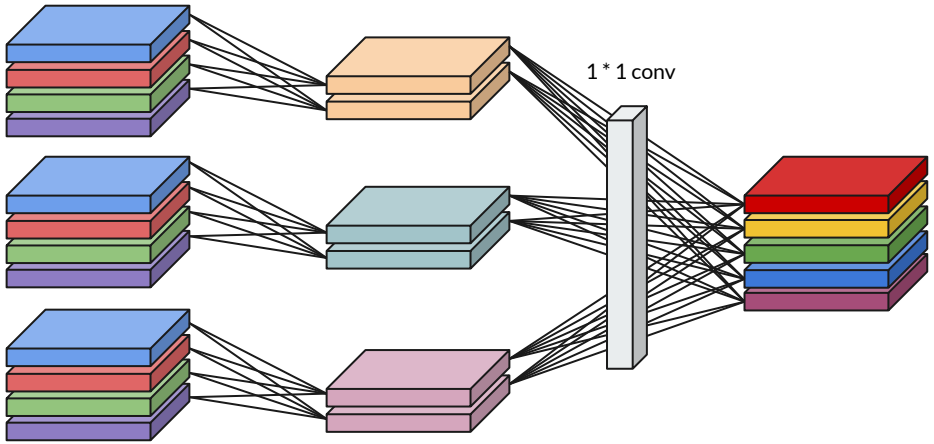
Looking Through Neighbors: I. Group Convolution



Looking Through Neighbors: II. 3D Convolution



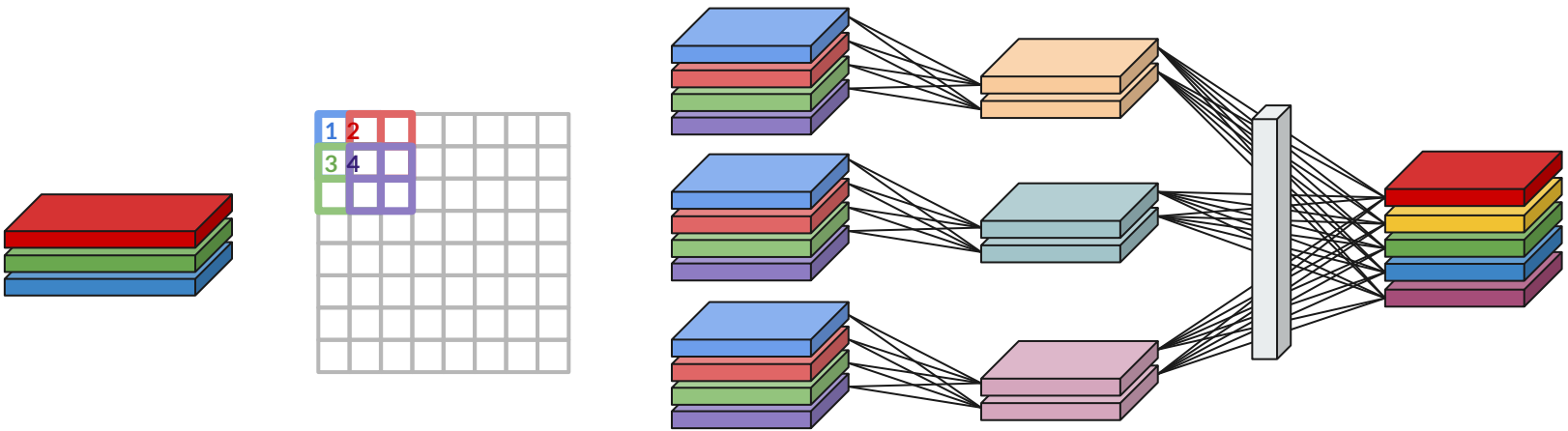
Aggregate by using 1 * 1 convolutions



C * 4 * H/2 * W/2

Cout * H/2 * W/2

Shift Invariant Module



Cin * H * W Offsetted Operations C * 4 * H/2 * W/2 Group/3D conv 1*1 conv Cout * H/2 * W/2

Experiment Results

[1] [1904.11486] Making Convolutional Networks Shift-Invariant Again (arxiv.org)
[3] convolutional neural networks - What do the numbers in this CNN architecture stand for? - Artificial Intelligence Stack Exchange

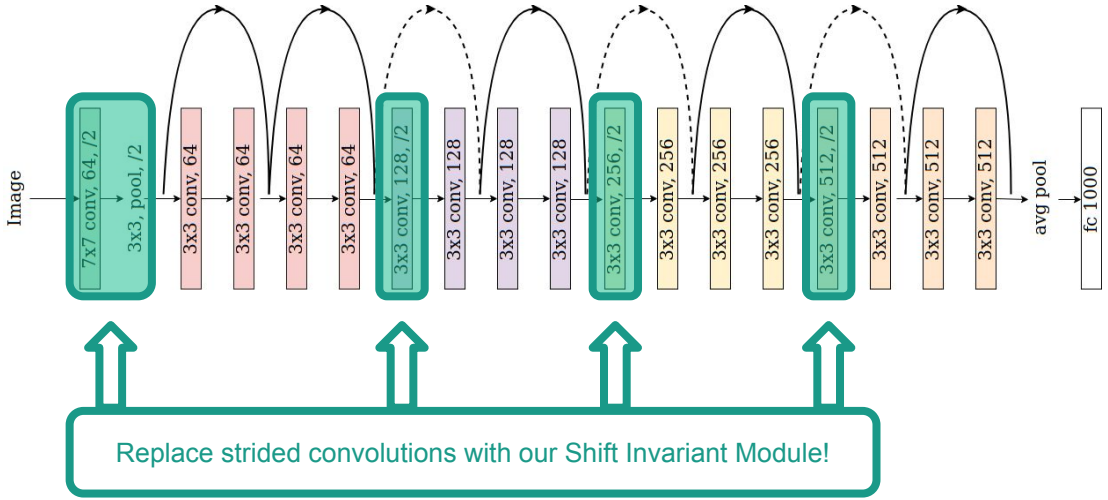
Settings

Task: Classification

Dataset: ImageNet

Model: RestNet18

Epochs: 16



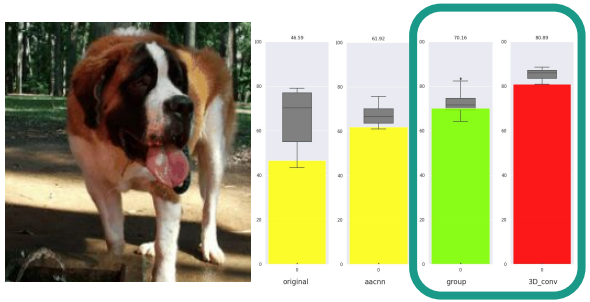
Metrics

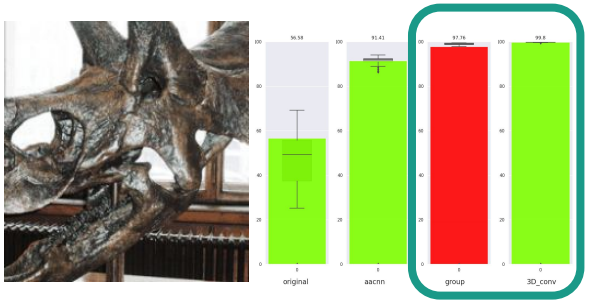
Top 1, Top 5 Accuracy

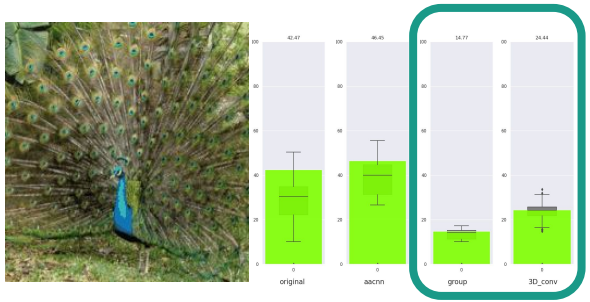
Classification Consistency:

$$E_{X,h1,w1,h2,w2}\mathbf{1}\{argmax\ P(Shift_{h1,w1}(X)) = argmax\ P(Shift_{h2,w2}(X))\}$$

Qualitative







Quantitative

Model	Consistency	Top 1	Top 5
Original	83.012	65.3699	87.0039
AACNN	87.834	**69.1259**	**89.0500**
Group	**90.422**	68.7780	88.7399
3D	90.200	68.5620	88.4079

** All higher the better*

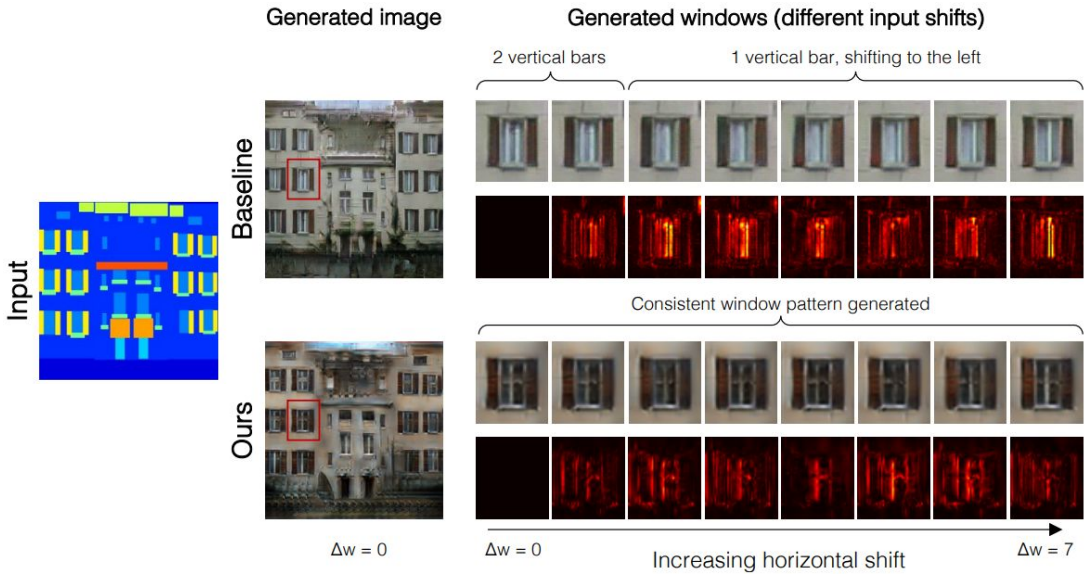
Future Work

[1] [1904.11486] Making Convolutional Networks Shift-Invariant Again (arxiv.org)

Other tasks

We want to try on other tasks!

- Segmentation
- Conditioned image generation

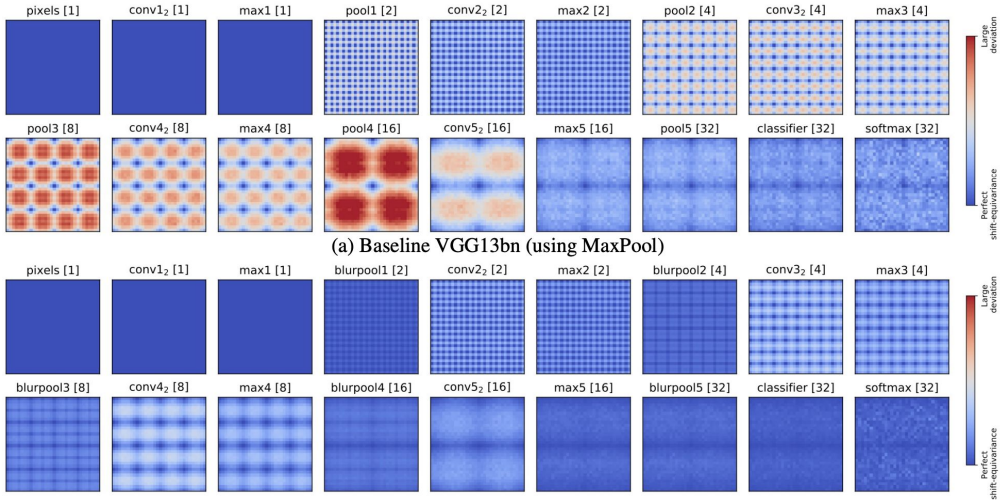


Visualization

We want to try on other tasks!

- Segmentation
- Conditioned image generation

Feature Visualization



Thanks!