Orthogonal Convolutional Neural Networks

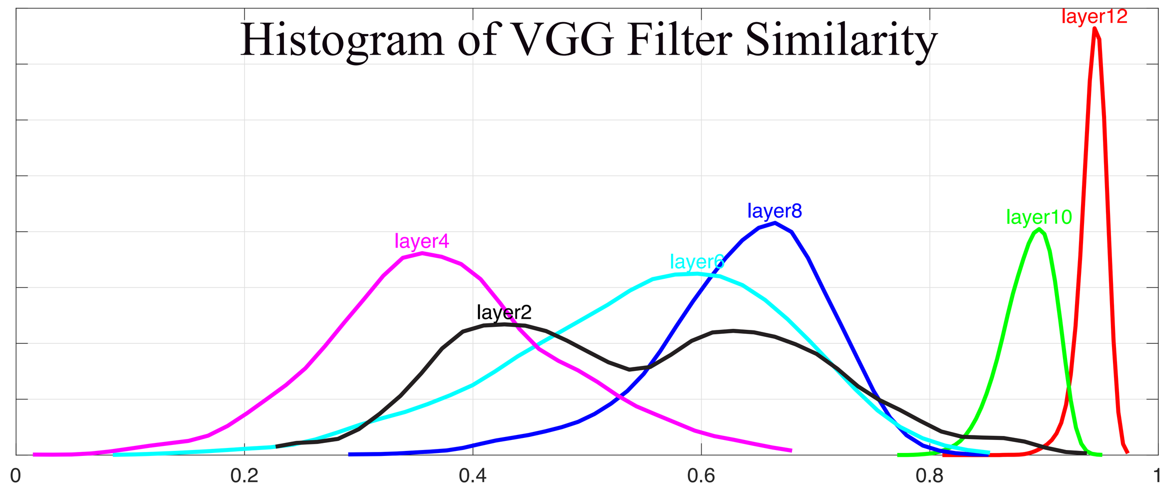
Jiayun (Peter) Wang   Yubei Chen   Rudrasis Chakraborty   Stella X. Yu

INTERNATIONAL COMPUTER SCIENCE INSTITUTE

REDWOOD CENTER for Theoretical Neuroscience

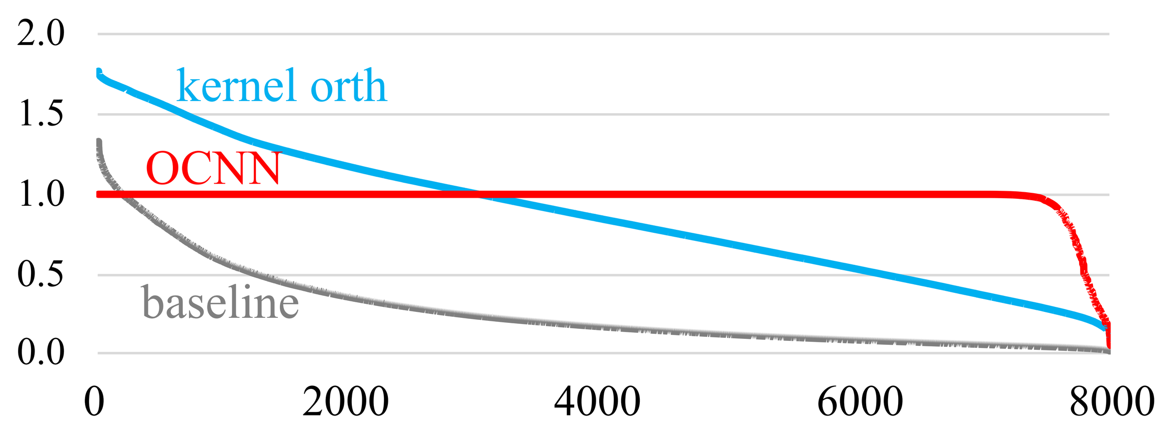
Motivations

- **Feature redundancy** in CNNs:
  - Features are similar
  - Network capacity is not fully utilized



- **Instability** of training CNNs:
  - Spectrum of weight matrix is highly imbalanced
  - Scaling power to different images is imbalanced

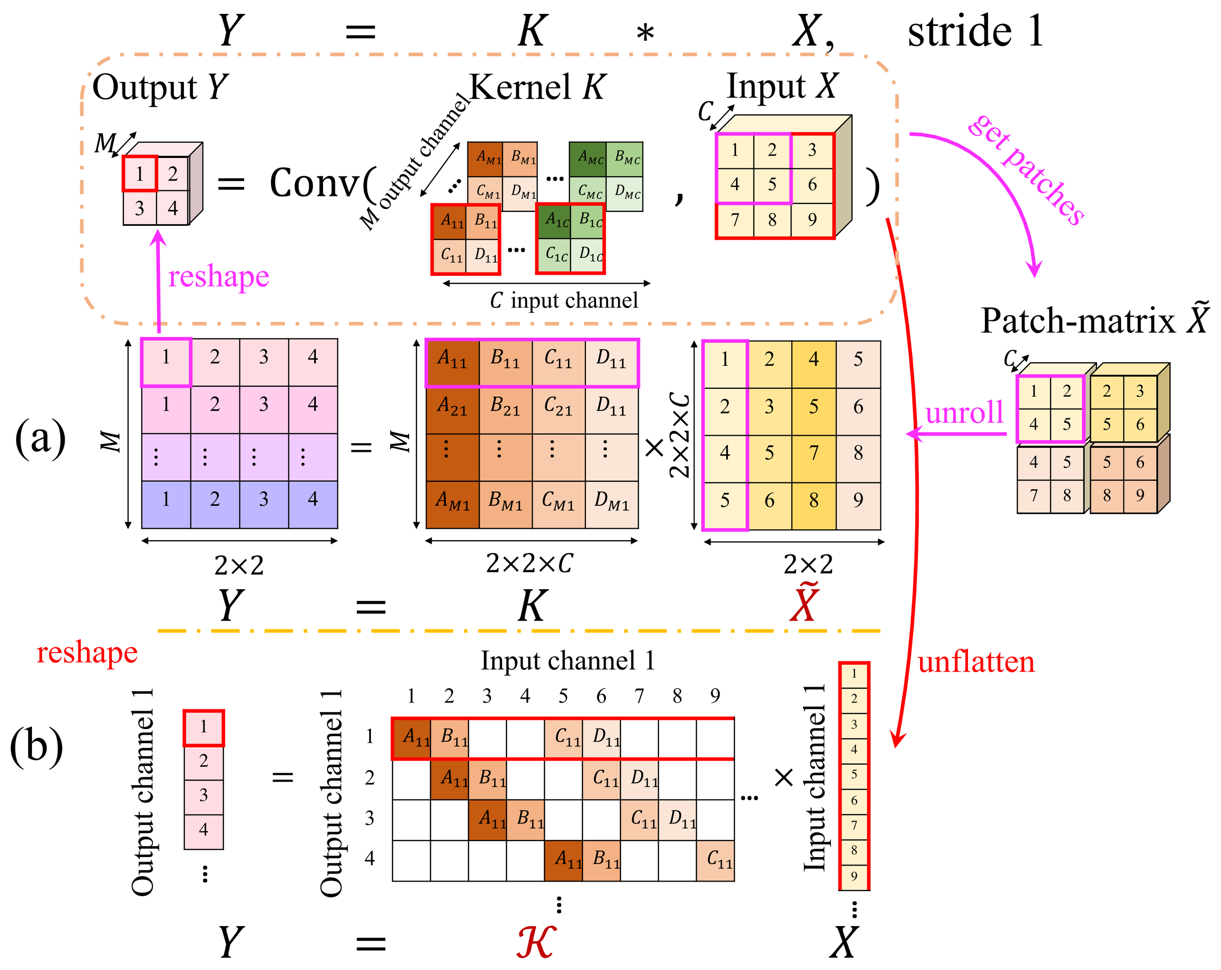
Spectrum of Kernel $\mathcal{K}$



- We propose to use **orthogonal convolutions** in CNNs
  - Reduce feature redundancy
  - Improve feature expressiveness
  - Enhance network stability
  - Increase robustness under attack

Convolution Layer Analysis

- A convolution $Y = \text{Conv}(K, X)$ can be considered as matrix multiplications in two formats:



(a) **Previous work** converts input X to patch-matrix $\widetilde{X}$; retains kernel K.

$$Y = K\widetilde{X}$$

(b) **Our work** converts K to a Toeplitz matrix $\mathcal{K}$; retains input X.

$$Y = \mathcal{K}X$$

- Our form **directly** analyzes the transformation between the input and the output.
- We further constrain convolutions to be orthogonal.

Orthogonal Convolutions

Row Orthogonality:

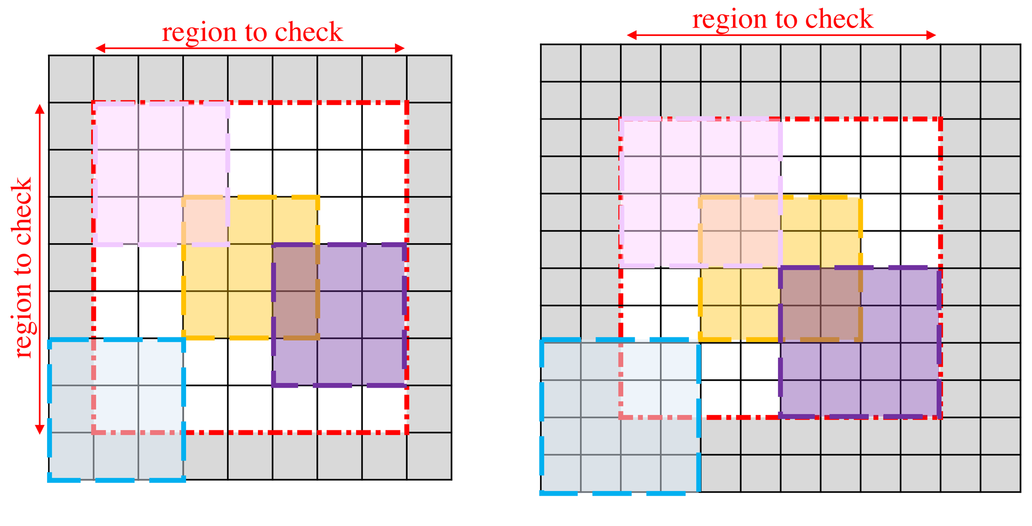
- Row orthogonal convolution can be achieved if:

$$\text{Conv}(K, K, \text{padding} = P, \text{stride} = S) = I_{r0}$$

$$P = \lfloor \frac{k-1}{S} \rfloor \cdot S.$$

$I_{r0} \in \mathbf{R}^{M \times M \times (2P/S+1) \times (2P/S+1)}$ has zeros entries except for the center $M \times M$ entries as an identity matrix.

- Intuitively, only checking overlapping regions is enough:



(a) kernel size=3, stride=1   (b) kernel size=4, stride=2

Column Orthogonality:

$$\mathcal{K}(\cdot, ihw) = \mathcal{K}\mathbf{e_{ihw}} = \text{Conv}(K, E_{i,h,w})$$

$$\langle \mathcal{K}(\cdot, ih_1w_1), \mathcal{K}(\cdot, jh_2w_2) \rangle = \begin{cases} 1, (i, h_1, w_1) = (j, h_2, w_2) \\ 0, \text{otherwise} \end{cases}$$

Comparisons:

- **Previous work**: Kernel orthogonality:

$$\begin{cases} KK^T = I \\ K^TK = I \end{cases}$$

- **Our work**: Convolutional orthogonality:

$$\begin{cases} \text{Conv}(K, K, \text{padding} = 0) = I_{r0} \\ \text{Conv}(K^T, K^T, \text{padding} = 0) = I_{c0} \end{cases}$$

Orthogonal CNNs:

- We add additional orthogonality loss to the final loss:

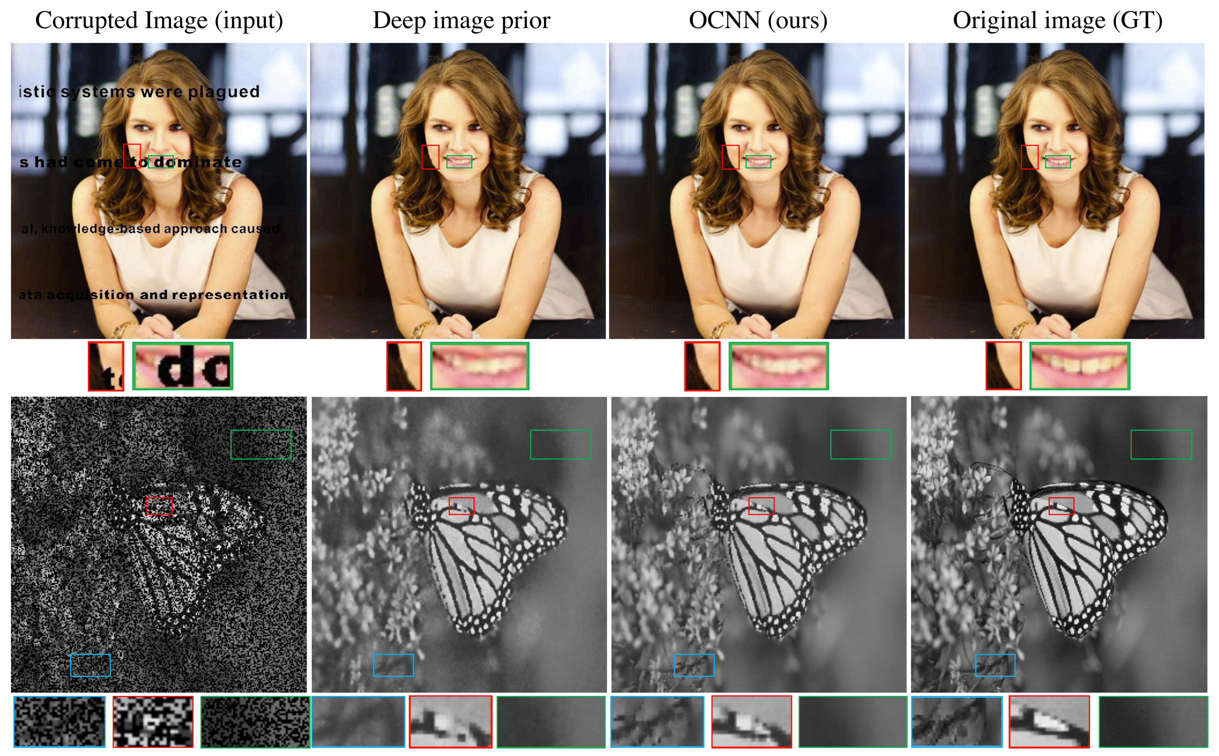
$$L = L_{\text{task}} + \lambda L_{\text{orth}}$$

$$L_{\text{orth}} = \|\text{Conv}(K, K) - I_{r0}\|_F^2$$

Experiments & Results

Image Inpainting (Unsupervised):

Improve low-level visual feature: 4.3 PSNR gain.



Fine-grained Image Retrieval:

Improve high-level visual feature: Trained on ImageNet only. 3% gain on top-5 classification.

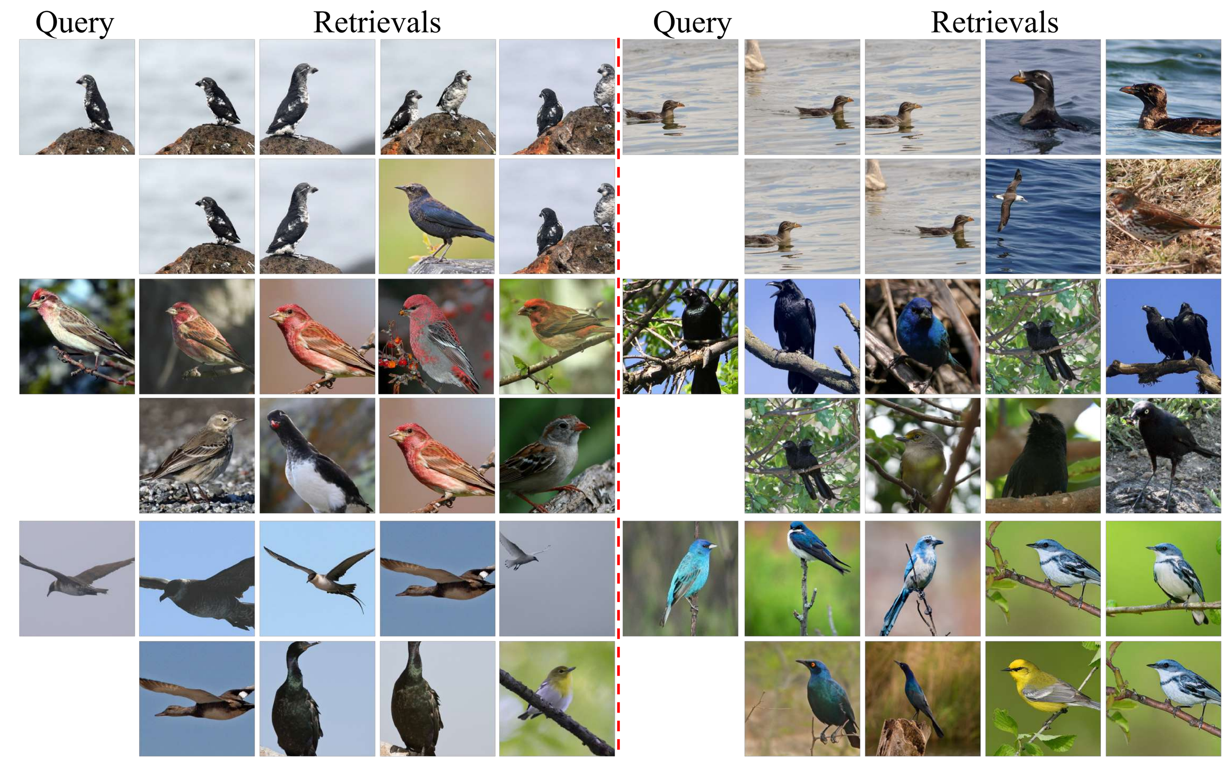
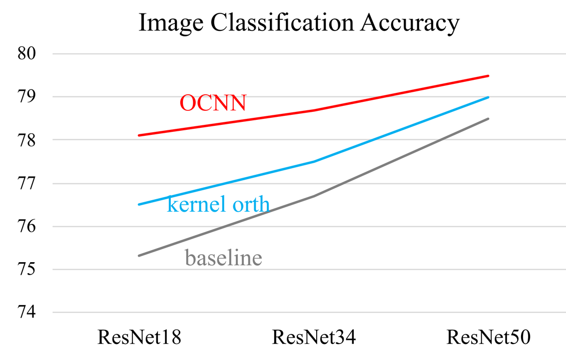


Image Classification:

3% gain on CIFAR100 and 1% gain on ImageNet.



(a) CIFAR100.

	Top-1 error	Top-5 error
ResNet34 (baseline) [23]	26.70	8.58
OMDSM [28]	/	9.68
kernel orthogonality [52]	/	8.43
SRIP [6]	/	8.32
OCNN (ours)	**25.87**	**7.89**

(b) ImageNet.

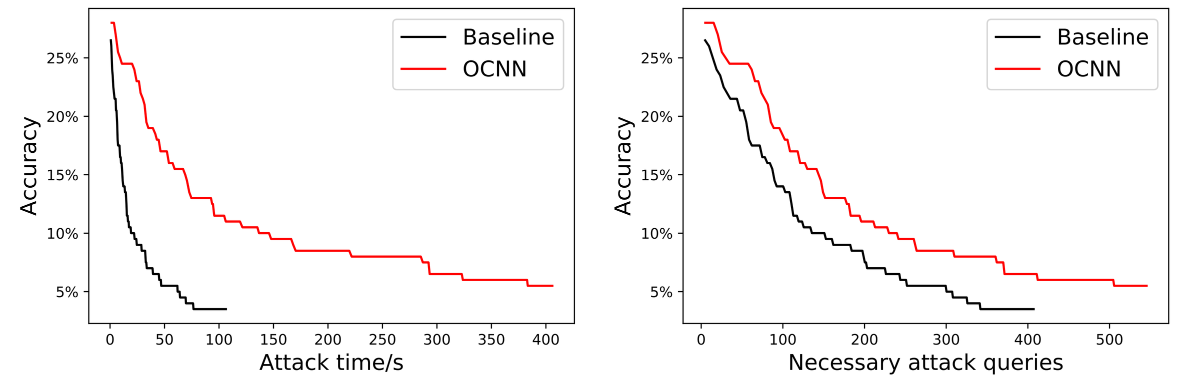
Semi-supervised Learning:

Consistent 2% - 3% gain under different fractions of labelled data on CIFAR100.

% of training data	10%	20%	40%	60%	80%	100%
ResNet18 [23]	31.2	47.9	60.9	66.6	69.1	75.3
kernel orthogonality [52]	33.7	50.5	63.0	68.8	70.9	76.5
Conv-orthogonality	**34.5**	**51.0**	**63.5**	**69.2**	**71.5**	**78.1**
Our gain	3.3	3.1	2.6	2.6	2.4	2.8

Robustness under Attack:

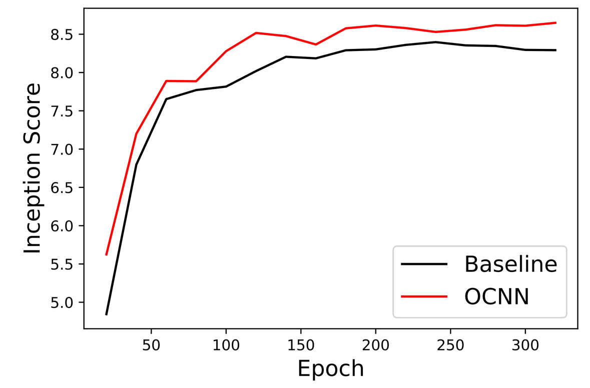
7x time and **1.7x attack queries** to attack model.



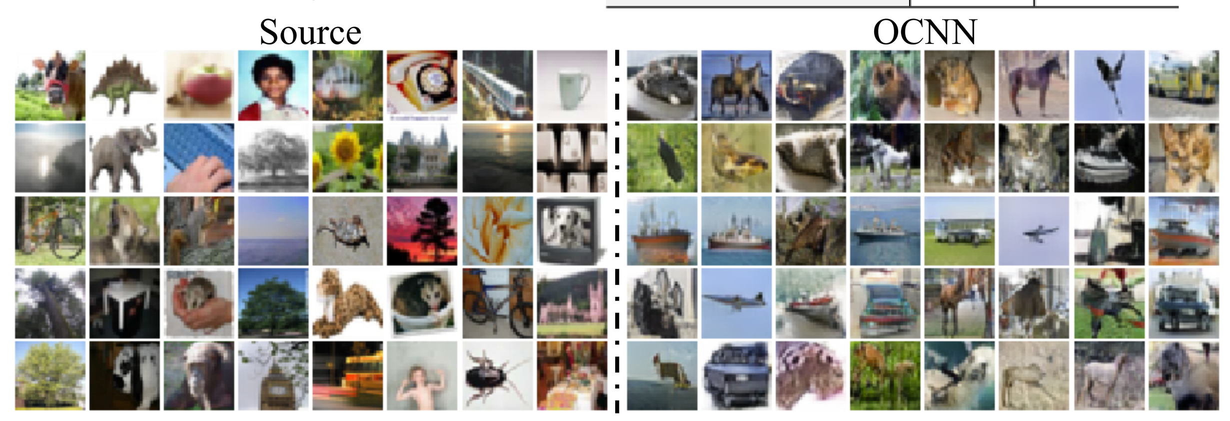
	Attack time/s	# necessary attack queries
ResNet18 [23]	19.3	27k
OCNN (ours)	136.7	46k

Image Generation:

- **Improve comprehensive visual feature.**
- 1.3 gain in FID.
- **Faster GAN convergence.**



	IS	FID
PixelCNN [48]	4.60	65.93
PixelIQN [38]	5.29	49.46
EBM [16]	6.78	38.20
SNGAN [37]	8.22	21.70
BigGAN [9]	**9.22**	14.73
AutoGAN [18]	8.32	13.01
OCNN (ours)	8.63	**11.75**

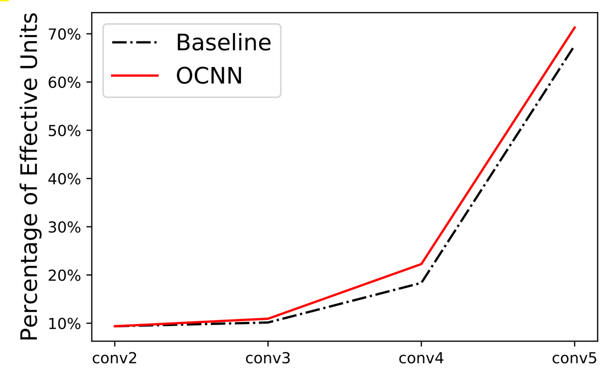


Space and time complexity:

Light-cost: 9% gain in training time only.
Same model size and test time.

Network Dissection:

- Network dissection analyzes how many unique detectors each layer have
- Ours has more detectors with more concepts.



Project Webpage:
Code & Paper & Model