# The Report about CirCNN

authors: Tianma Shen

University of Shanghai for Science and Technology

December 8, 2017

# Outline

# Outline

# Background

- Large-scale deep neural networks (DNNs)
- Limitations of computer performance
- Difficult tasks with big data

# Motivation

- Reduce weight storage (model size)
- Accelerate the computation
- Maintain accuracy

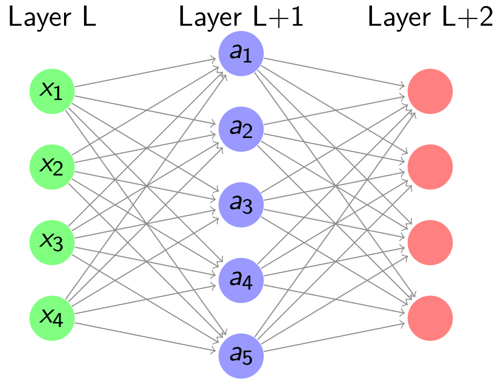
# Outline

# Novelty of CirCNN

- Supporting both FC and CONV layers
- Block-circulant matrices

# Outline

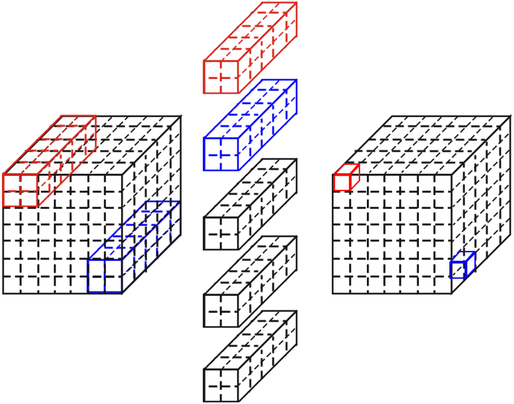
# Full Connect Layers




$$a_1 = f(W_{11} \cdot X_1 + W_{12} \cdot X_2 + W_{13} \cdot X_3 + W_{14} \cdot X_4) \quad (1)$$

$$a_5 = f(W_{51} \cdot X_1 + W_{52} \cdot X_2 + W_{53} \cdot X_3 + W_{54} \cdot X_4) \quad (2)$$

# Convolution Layer



$$Y(x,y,p)=\sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{c=1}^{C}F(i,j,c,p)\,X(x+i-1,y+j-1,c) \quad (3)$$

# Circulant matrices

$$\begin{bmatrix} W_{11} & W_{12} & W_{13} & \dots & W_{1n-2} & W_{1n-1} & W_{1n} \\ W_{1n} & W_{11} & W_{12} & W_{13} & \dots & W_{1n-2} & W_{1n-1} \\ W_{1n-1} & W_{1n} & W_{11} & W_{12} & W_{13} & \dots & W_{1n-2} \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ W_{13} & \dots & W_{1n-2} & W_{1n-1} & W_{1n} & W_{11} & W_{12} \\ W_{12} & W_{13} & \dots & W_{1n-2} & W_{1n-1} & W_{1n} & W_{11} \end{bmatrix} \tag{4}$$

# Discrete Fourier Transform

The discrete Fourier transform transforms a sequence of N numbers $x_0, x_1, \ldots, x_{N-1}$ into another sequence of complex numbers, $X_0, X_1, \ldots, X_{N-1}$ ,which is defined by

$$
\begin{aligned}
X_k &= \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi k \frac{n}{N}} &\quad (5)\\
&= \sum_{n=0}^{N-1} x_n \left[\cos\left(2\pi k \frac{n}{N}\right) - i \cdot \sin\left(2\pi k \frac{n}{N}\right)\right] &\quad (6)
\end{aligned}
$$

where the last expression follows from the first one by Euler's formula.

The transform is sometimes denoted by the symbol $\mathcal{F}$, as in $\mathbf{X} = \mathcal{F}\{\mathbf{x}\}$or $\mathcal{F}(\mathbf{x})$ or $\mathcal{F}\mathbf{x}$.

# Convolution theorem

Let $\mathcal{F}$ denote the Fourier transform operator, so $\mathcal{F}\{f\}$ and $\mathcal{F}\{g\}$ are the Fourier transforms of $f$ and $g$, respectively. Then

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\} \tag{7}$$

$$f * g = \mathcal{F}^{-1}\{\mathcal{F}\{f\} \cdot \mathcal{F}\{g\}\} \tag{8}$$

where $\cdot$ denotes point-wise multiplication.
$*$ denotes convolution.
$\mathcal{F}^{-1}$ denotes inverse Fourier transform.

# Outline

# Forward propagation process

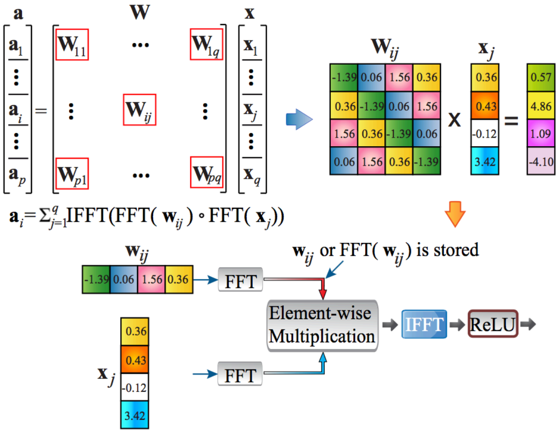



Figure: Illustration of the calculation of $W_x$ in the inference process

# Block-circulant matrices

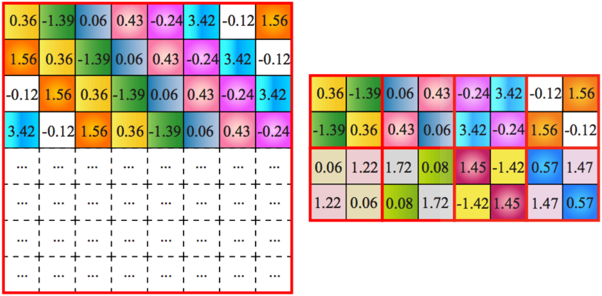



Figure: when the numbers of inputs and outputs are not equal

# Result

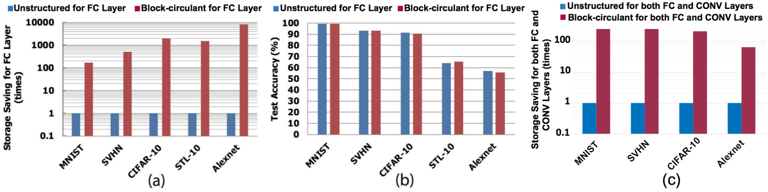



Figure: (a) Storage saving and (b) test accuracy after using block-circulant FC layer for DCNN models on di erent datasets. (c) Storage saving after using both block-circulant FC layer and block-circulant CONV layer for DCNNs on MNIST, SVHN, CIFAR-10, and ImageNet datasets.