

Scale Invariant Recognition by Weight Shared CNNs in Parallel

Ryo Takahashi

takahashi@ai.cs.kobe-u.ac.jp

Kobe University, Japan

Outline

- ▶ We introduce a novel **scale invariant CNNs** called “weight-shared multi-stage” network (WSMS-Net).
- ▶ Scale invariance enables the CNNs to acquire **higher performance** of image classification.

Index

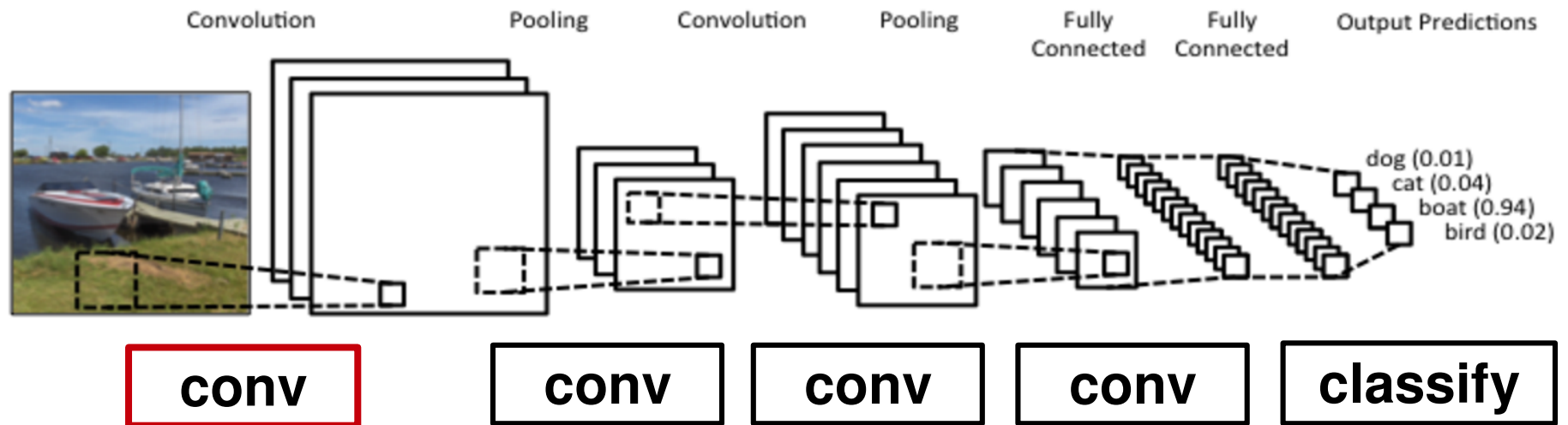
- ▶ **Introduction**
- ▶ **Our Method**
- ▶ **Experimental Results**
- ▶ **Future Works**

Index

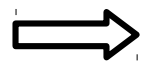
- ▶ **Introduction (CNN, Scale Invariance)**
- ▶ Our Method
- ▶ Experimental Results
- ▶ Future Works

What is the CNN?

CNN (Convolutional Neural Network)



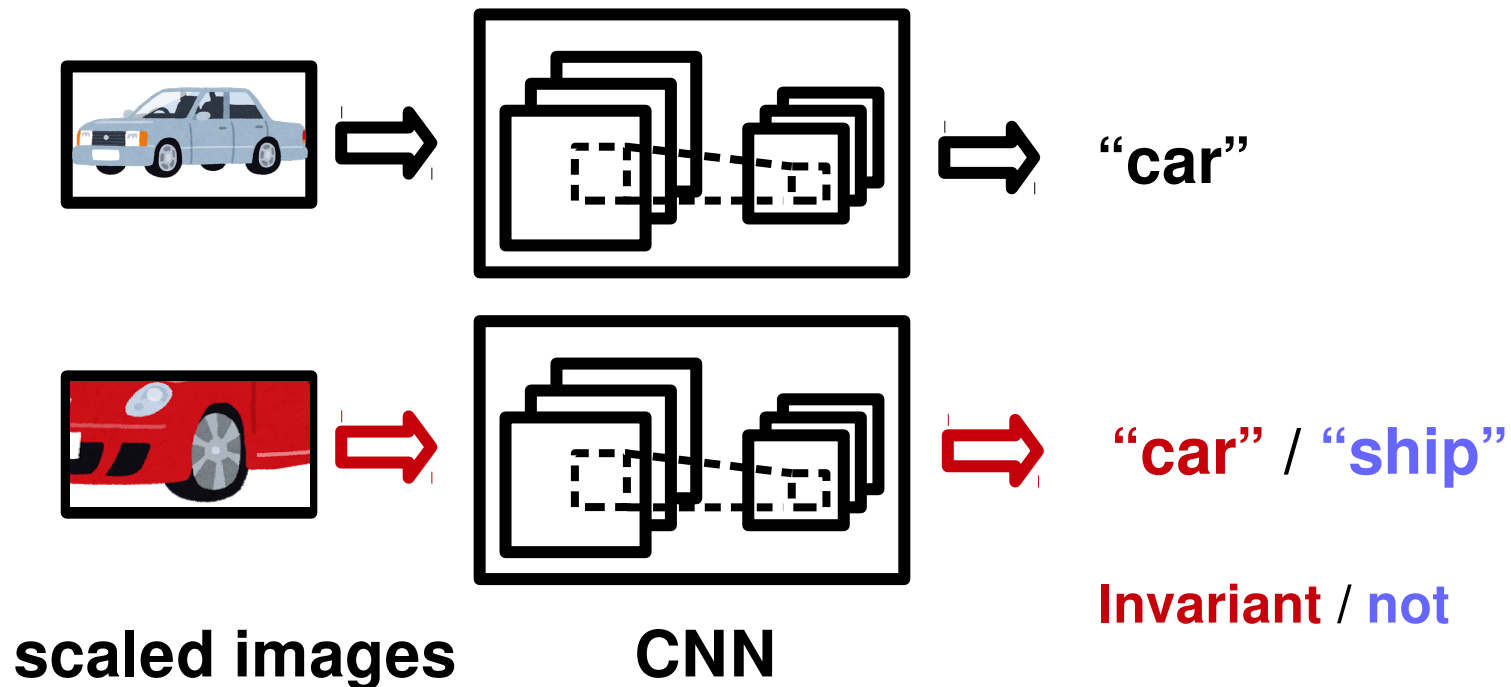
$$y_{ijk} = \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} w_{mnmk} x_{(i+m)(j+n)k} + b_k$$



w : weight parameters of neural network

What is the Scale Invariance?

- ▶ The “scale invariance” means that the CNN can classify the scaled images correctly.

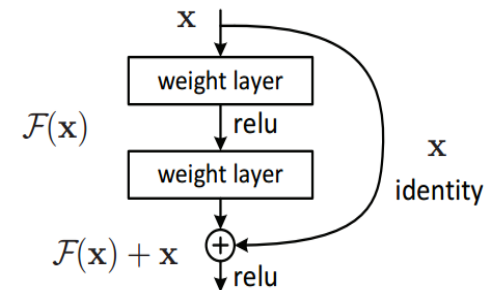


Why the Scale Invariance ?

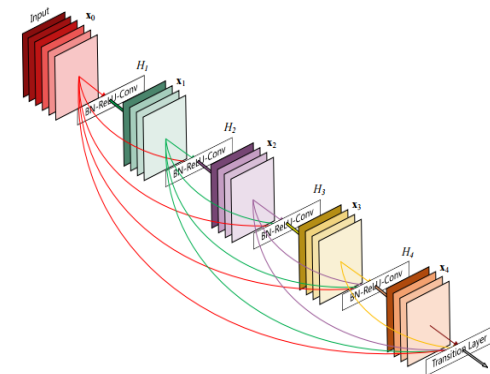
- ▶ Very deep CNNs of ResNet family achieved higher and higher accuracy on the image classification.

□ Examples of ResNet family CNNs

- **ResNet** [He+, 2015]
 - Shortcut connection
 - Ultra deep CNNs -100 layers or more-
- **DenseNet** [Zhuang+, 2016]
 - Densely connection of network
 - Ultra deep CNNs -100 layers or more-

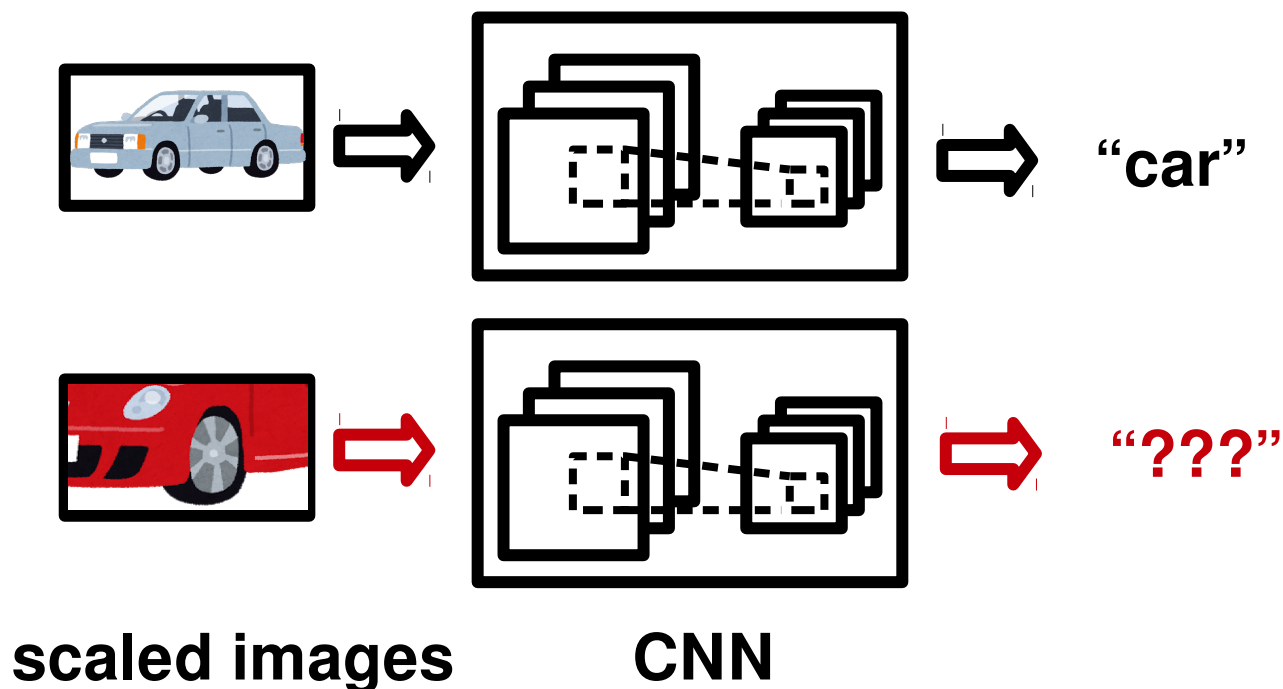


Residual block



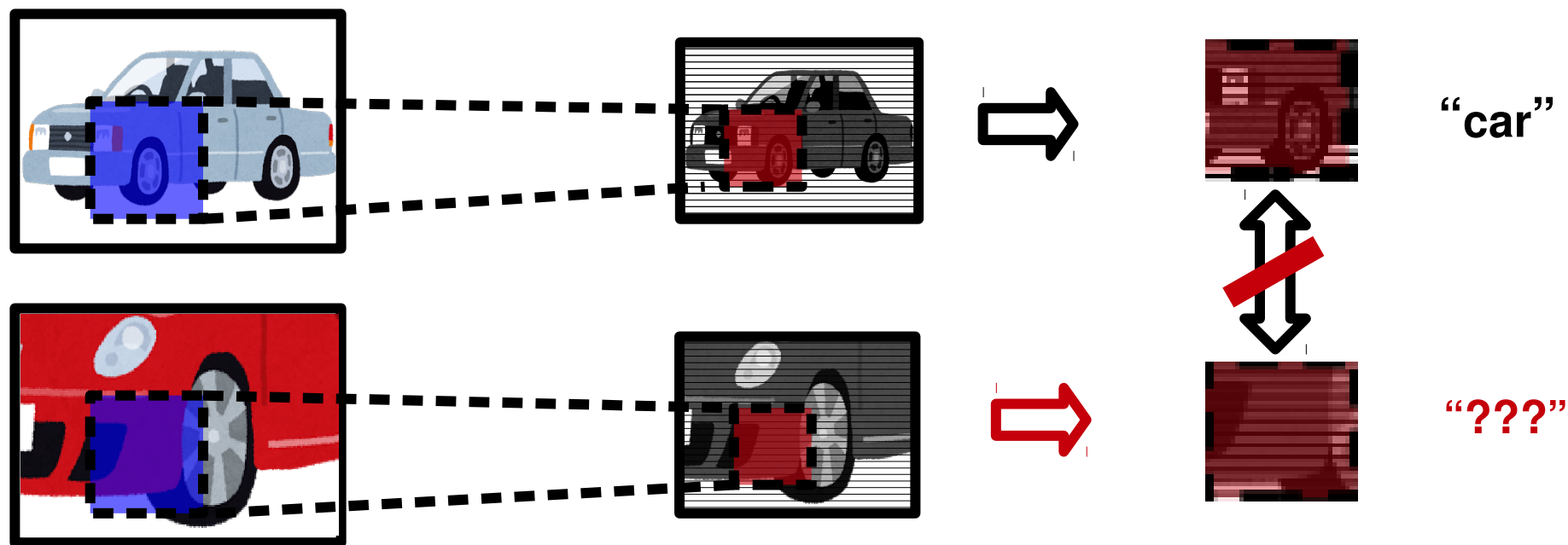
Why the Scale Invariance ?

- ▶ Very deep CNNs of ResNet family achieved higher and higher accuracy on the image classification.
- ▶ However, deepening the network cannot achieve the scale invariance because of the using convolution.



Why the Scale Invariance ?

- ▶ **Receptive field** of CNN does not cover the same area



inputs

feature maps

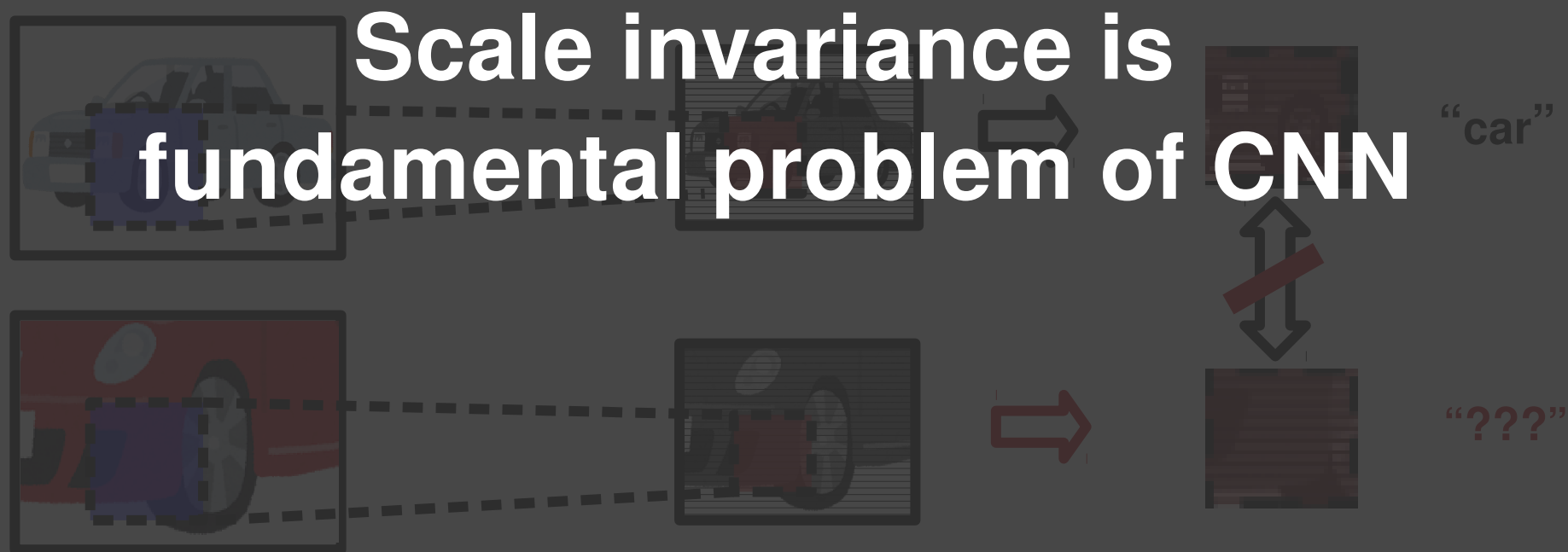
local feature

 receptive field

 output of CNN

Why the Scale Invariance ?

- ▶ **Receptive field** of CNN does not cover the same area



inputs

feature maps

local feature



receptive field



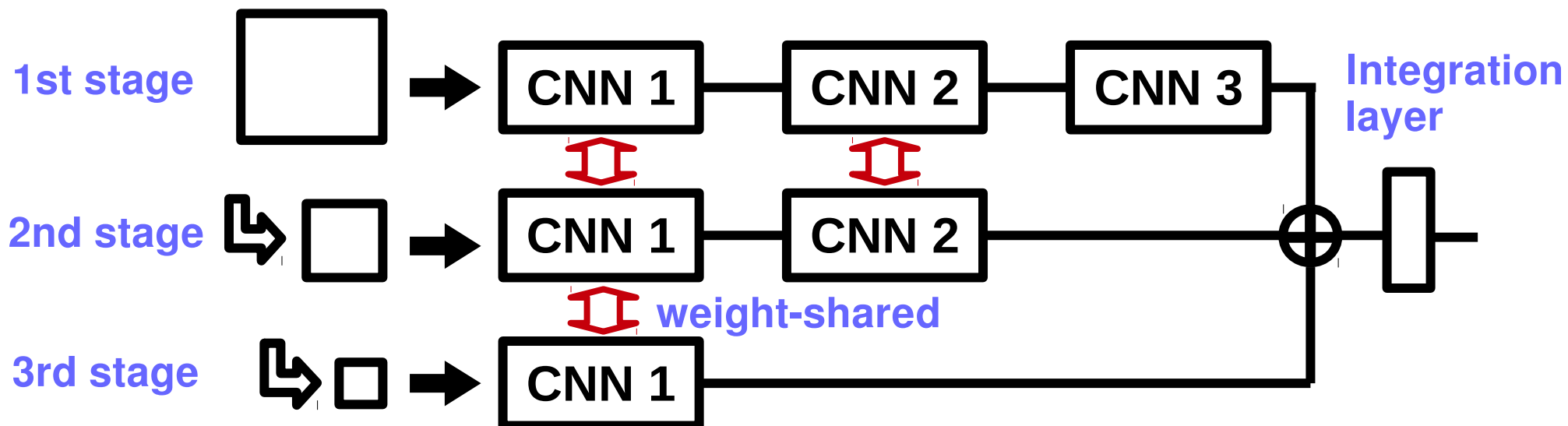
output of CNN

Index

- ▶ Introduction
- ▶ **Our Method (WSMS-Net)**
- ▶ Experimental Results
- ▶ Future Works

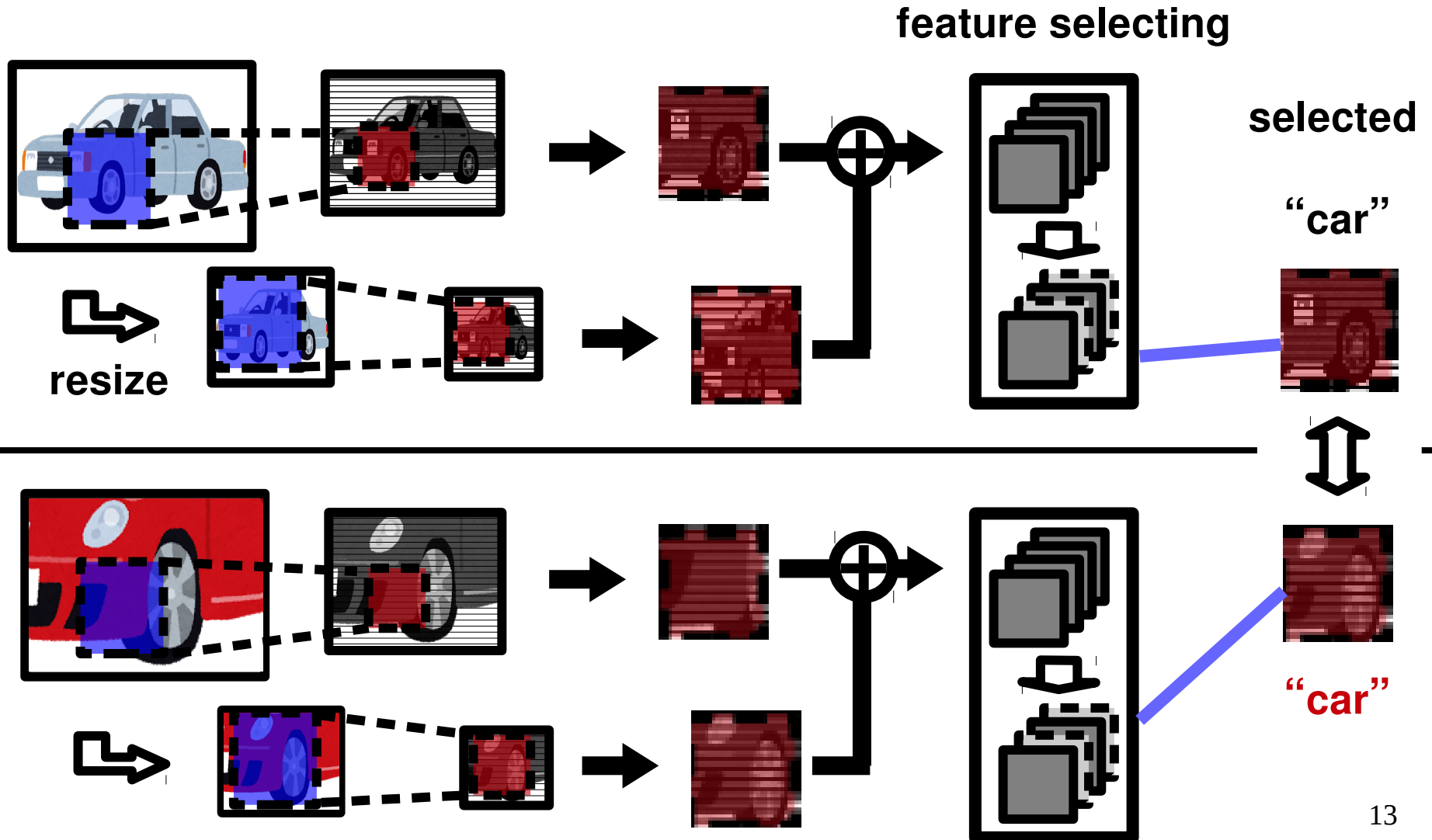
WSMS-Net for Scale Invariance

- ▶ WSMS-Net : Weight-shared multi-stage network
 - **Multi-stage** architecture for multi-scale feature acquiring
 - **Weight-shared** between multiple stages
 - **Integration** layer for **selecting feature** from multiple stages



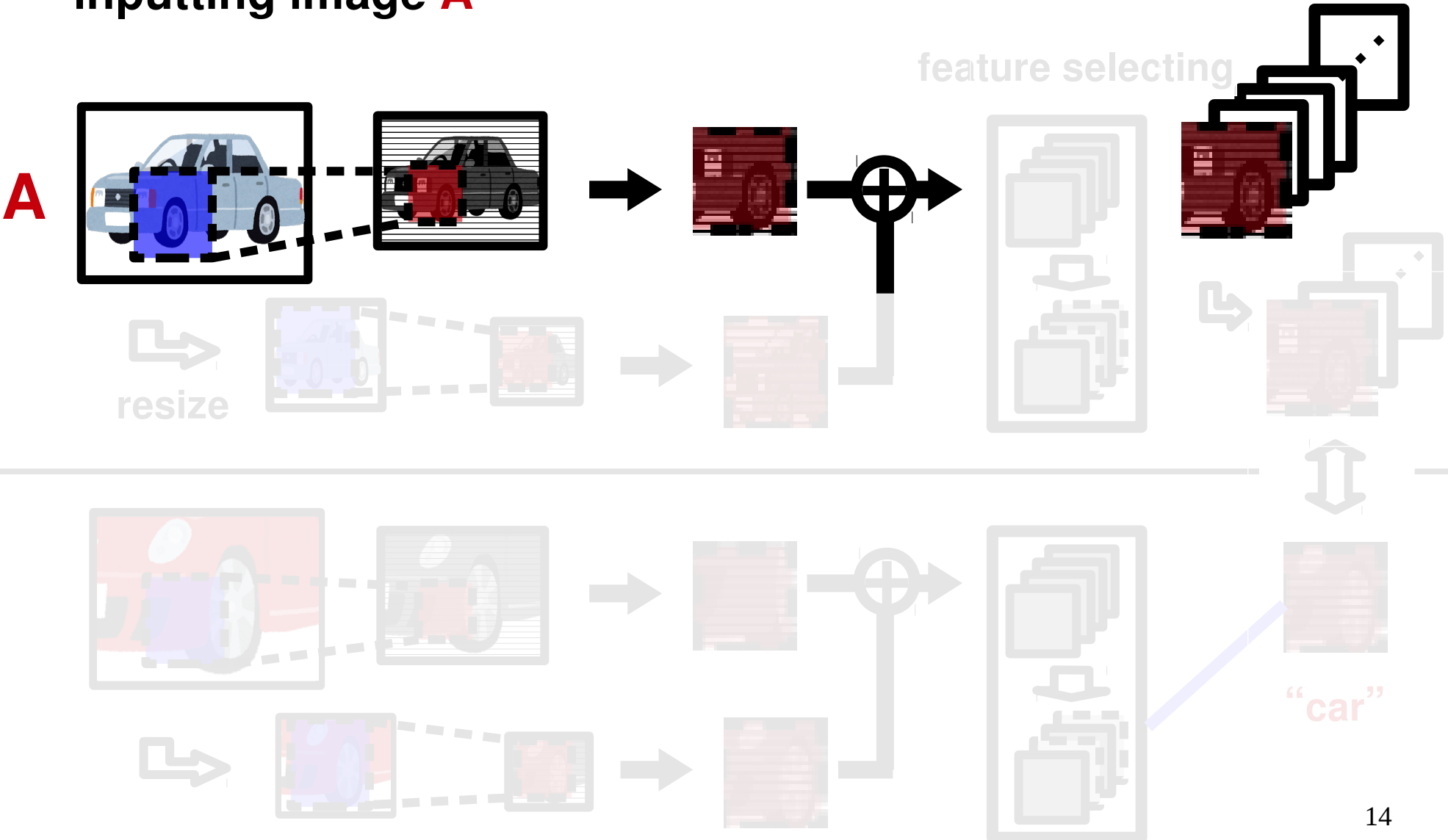
How is the Scale Invariance ?

overview



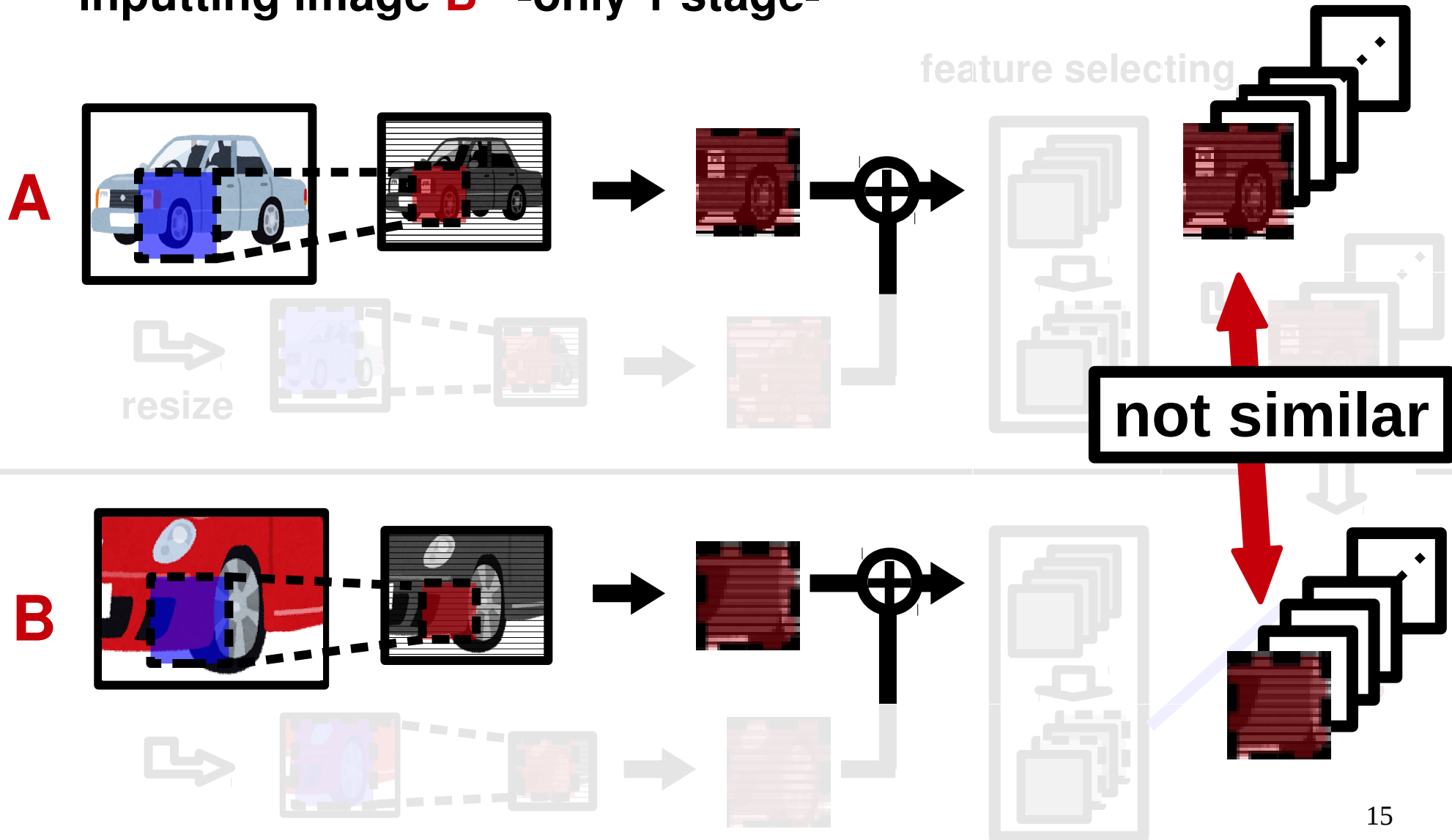
How is the Scale Invariance ?

inputting image **A**



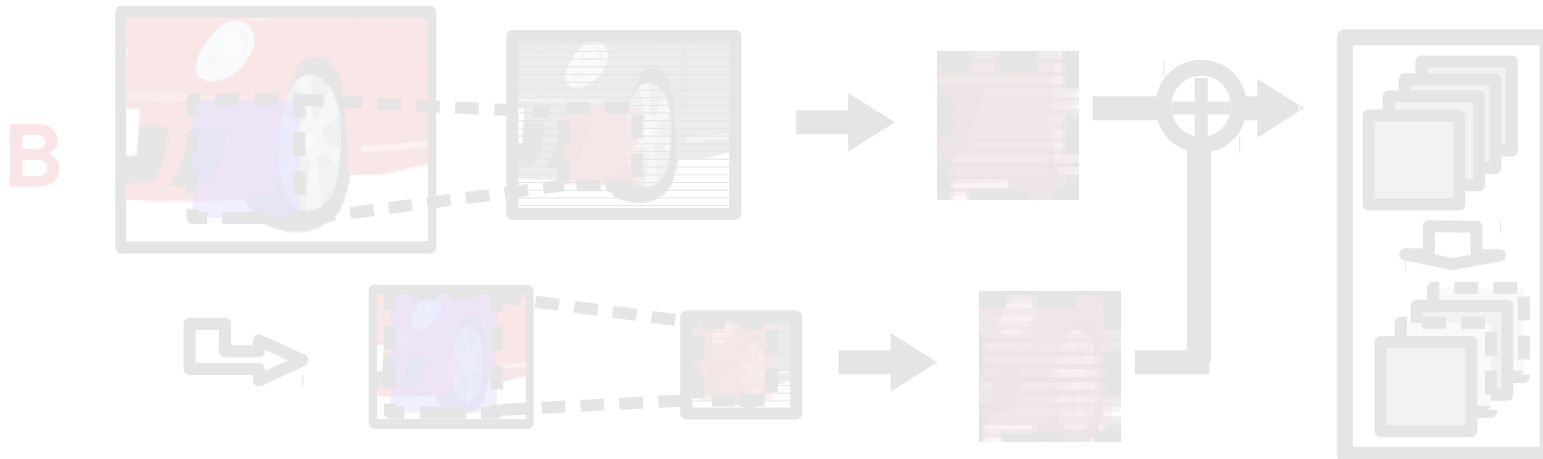
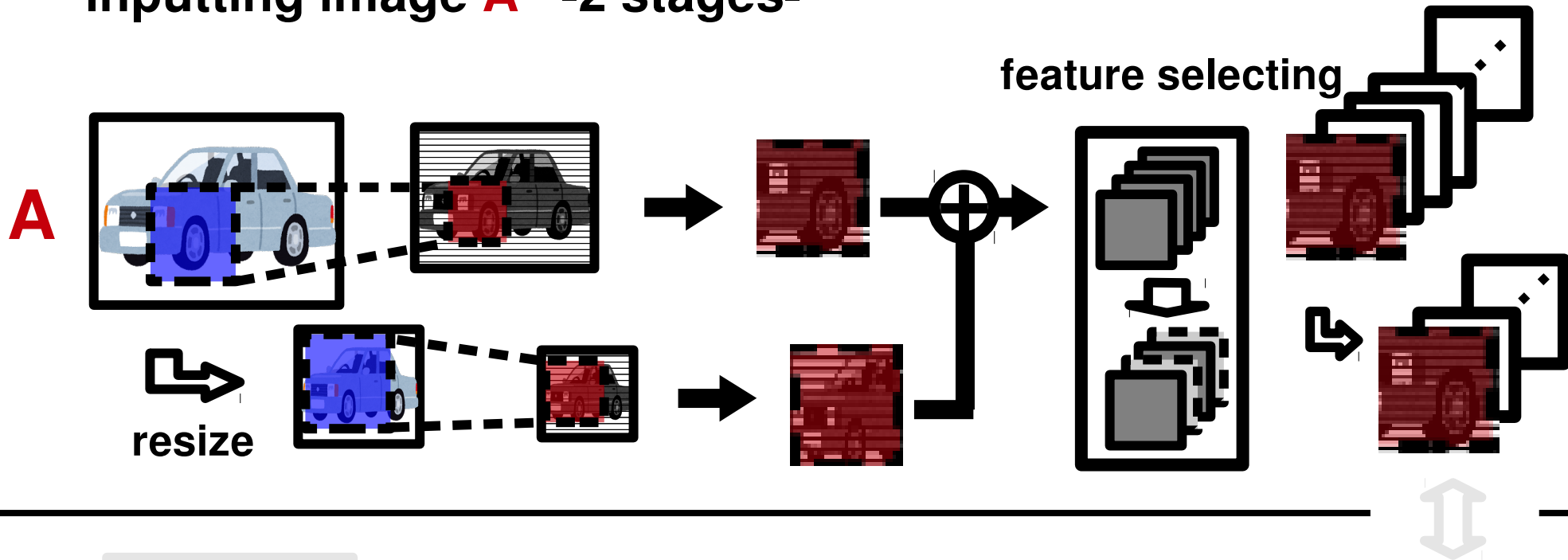
How is the Scale Invariance ?

inputting image **B** -only 1 stage-



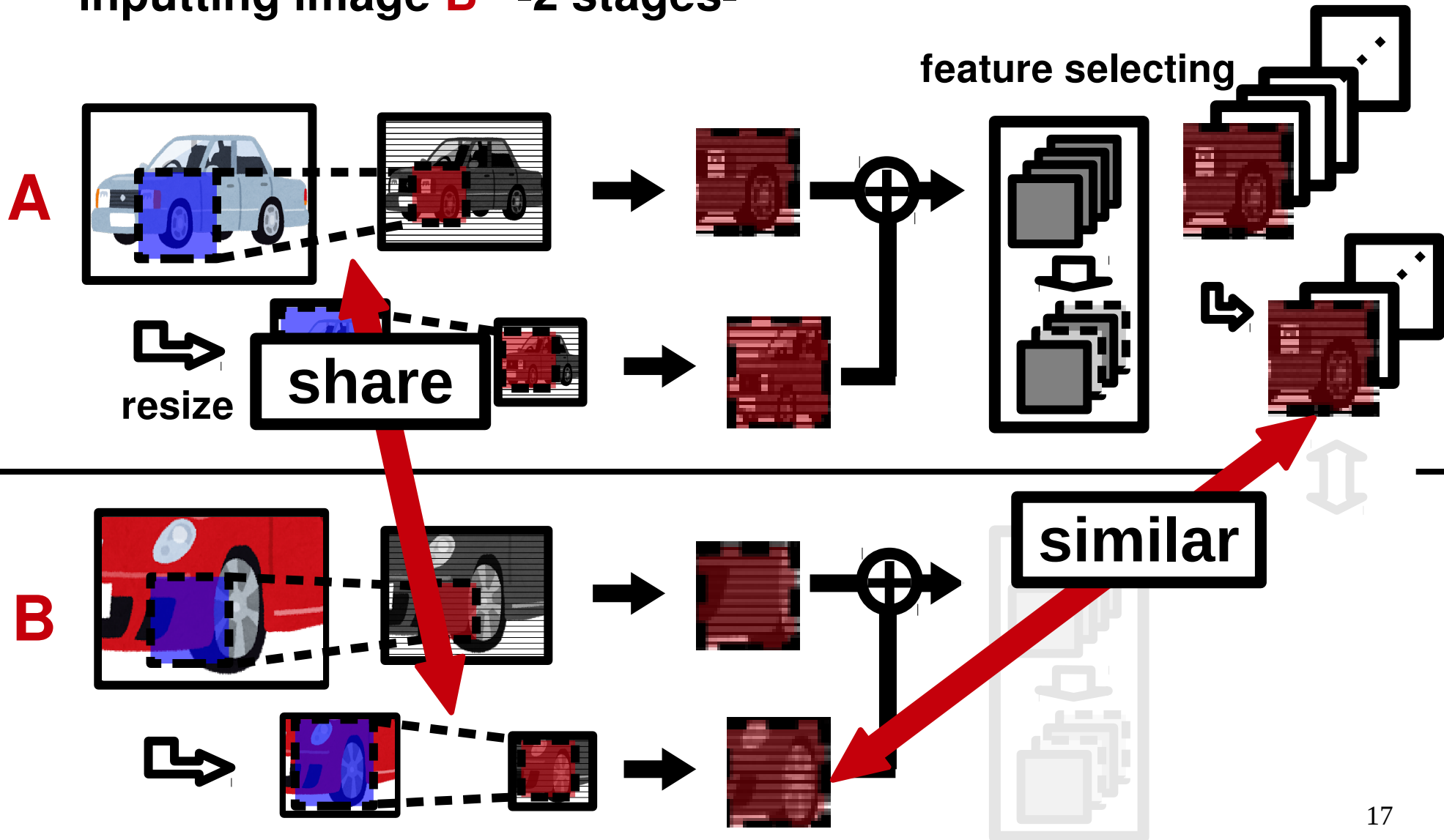
How is the Scale Invariance ?

inputting image **A** -2 stages-



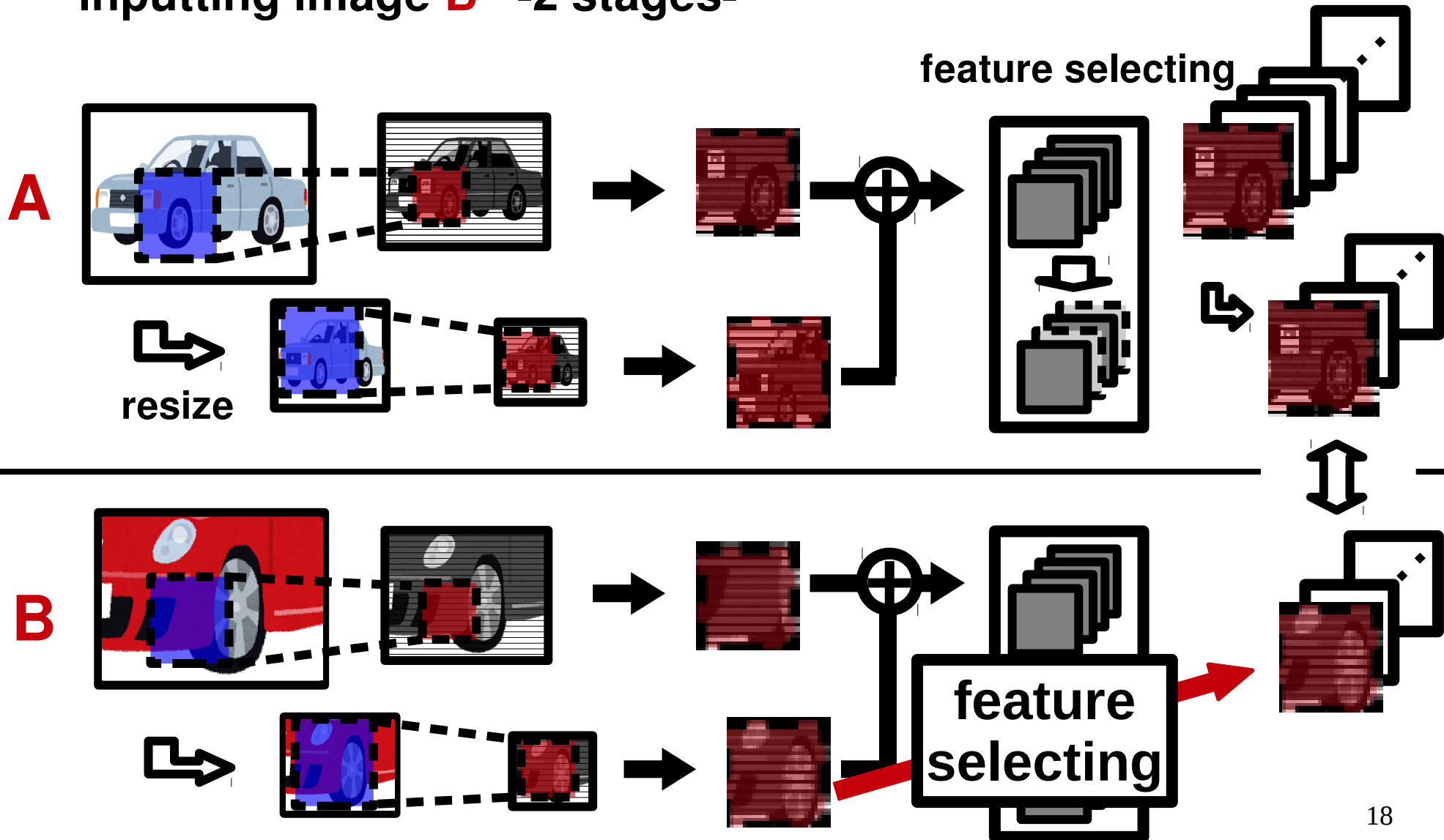
How is the Scale Invariance ?

inputting image **B** -2 stages-



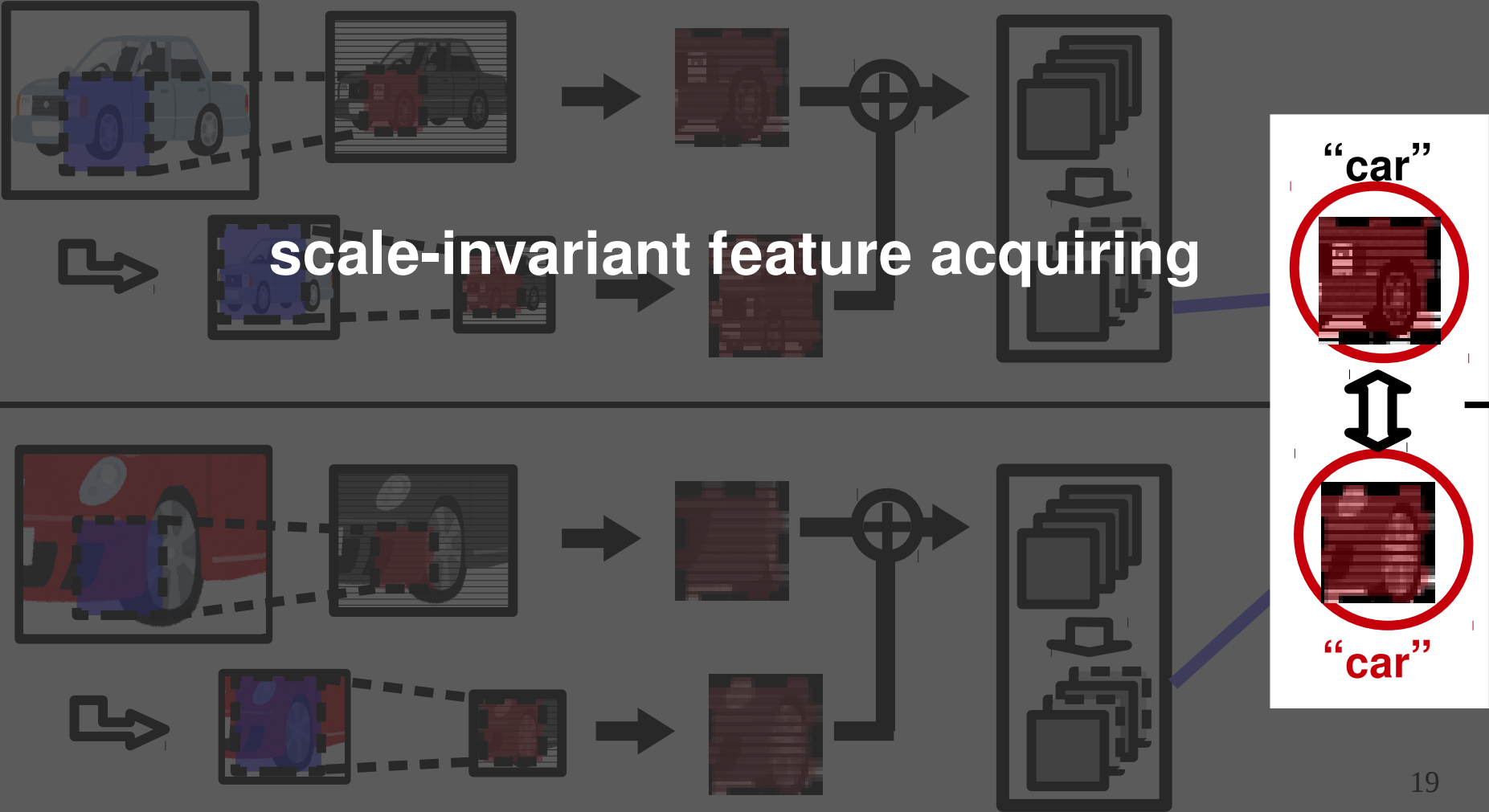
How is the Scale Invariance ?

inputting image **B** -2 stages-



How is the Scale Invariance ?

overview

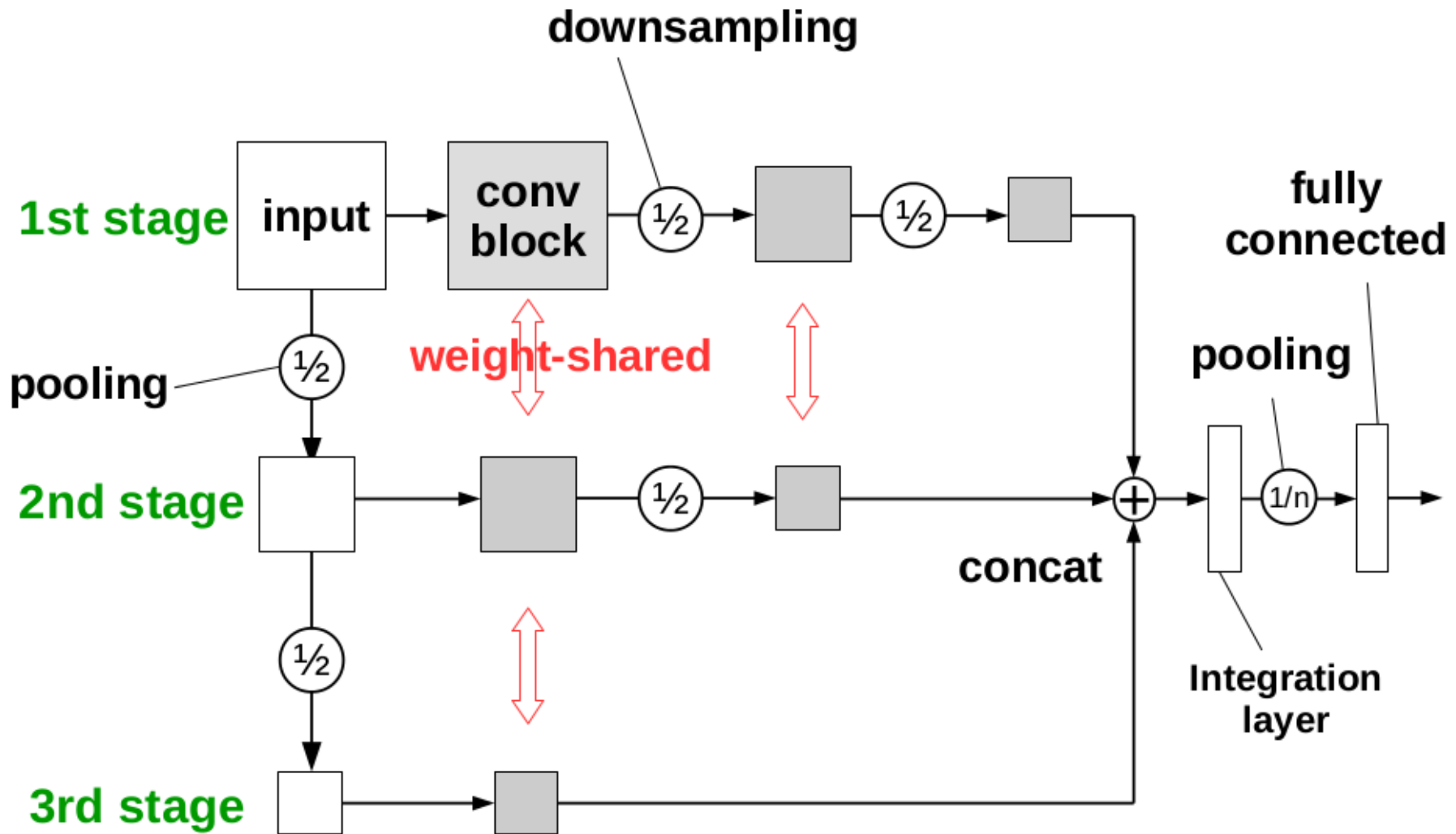


Index

- ▶ Introduction
- ▶ Our Method
- ▶ **Experimental Results (graduation thesis)**
- ▶ Future Works

Experimental Results -network-

► Basic architecture of WSMS-Net with 3 stages



Experimental Results -dataset-

► Datasets

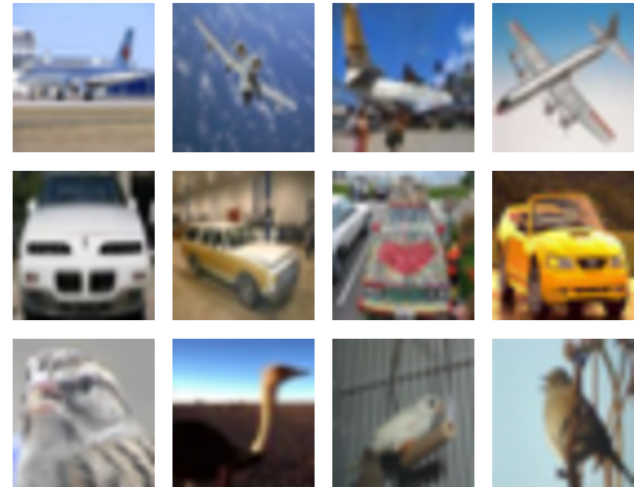
▪ CIFAR-10

10 classes color images

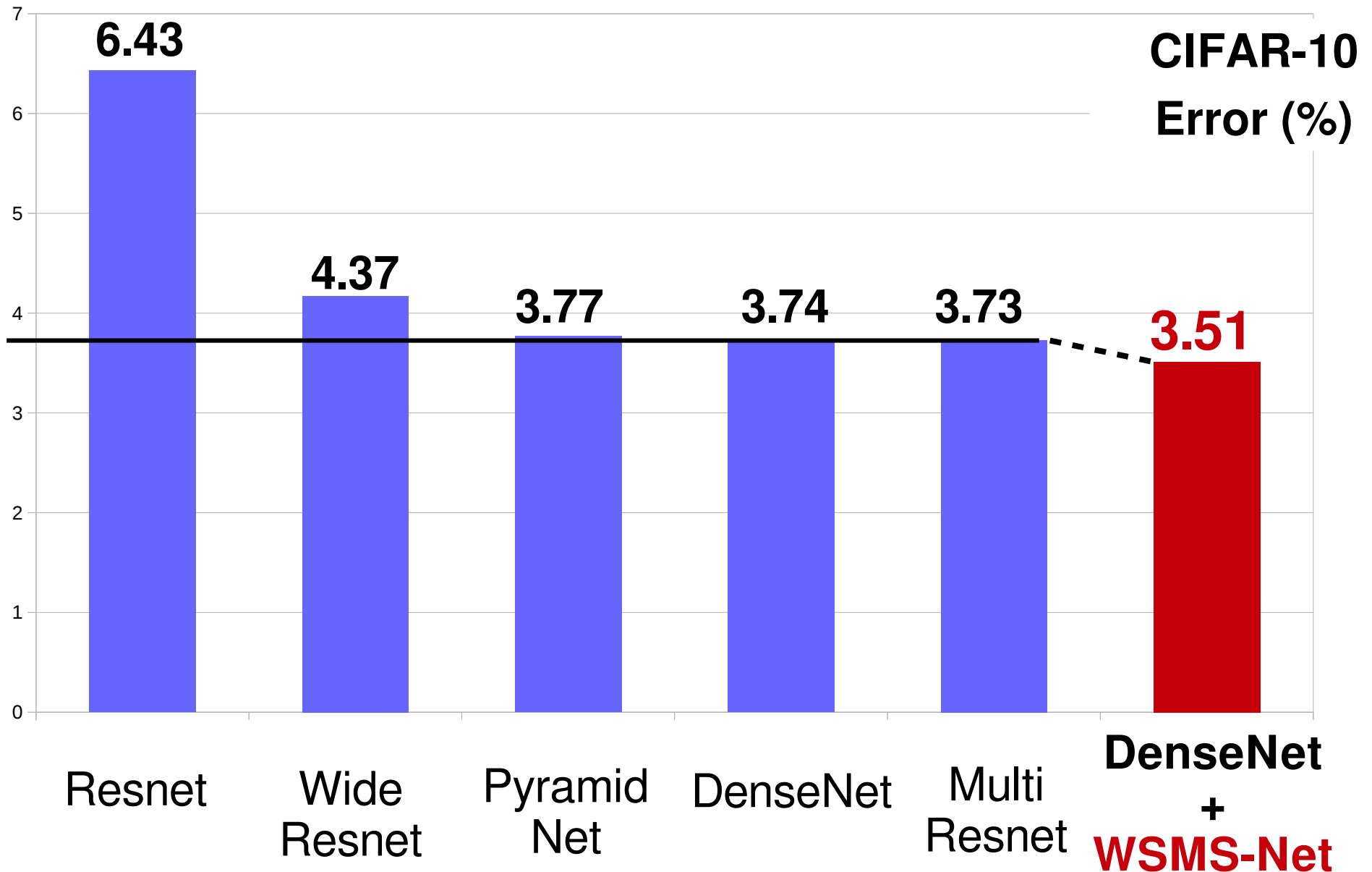
image size : 32 x 32 x 3

training data : 50,000 images

test data : 10,000 images

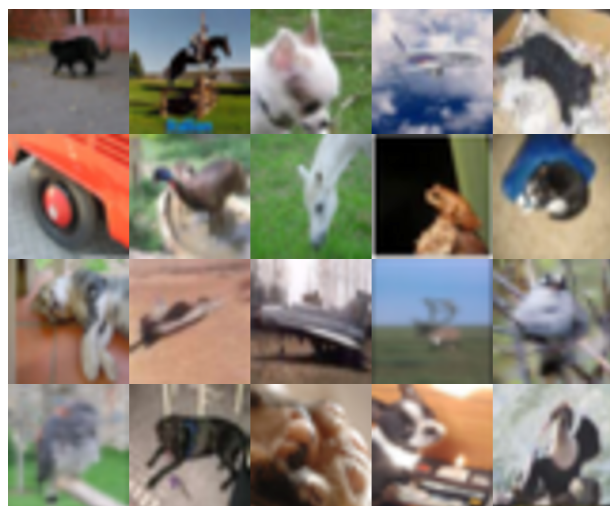


Experimental Results -accuracy-

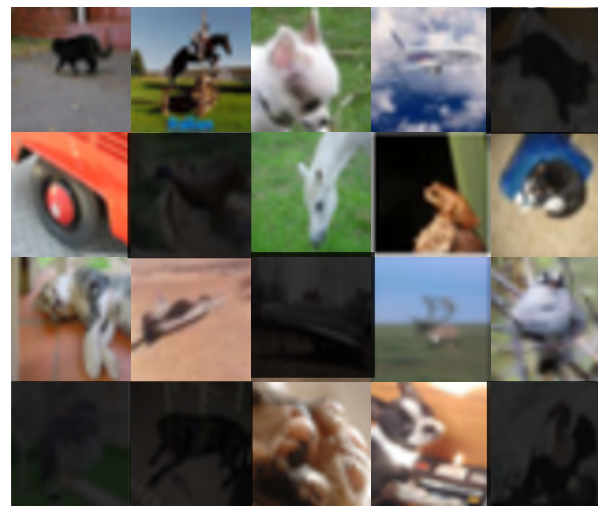


Experimental Results -visualize-

**WSMS-Net
newly
classify**



only scaled images



CIFAR-10 images

Index

- ▶ Introduction
- ▶ Our Method
- ▶ **Experimental Results (new)**
- ▶ Future Works

Index

▶ Introduction

▶ Our Method

▶ **Experimental Results (new)**

- **generalization capability (model, dataset)**
- **number of stages**
- **weight sharing**

Index

▶ Introduction

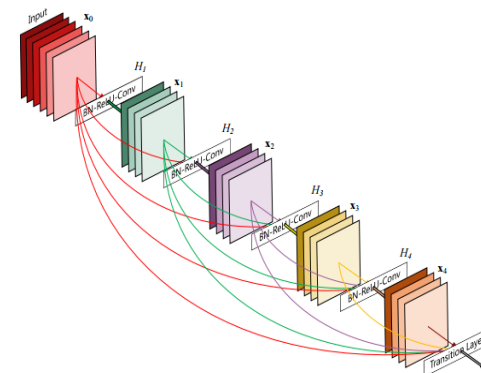
▶ Our Method

▶ **Experimental Results (new)**

- **generalization capability (model, dataset)**
- **number of stages**
- **weight sharing**

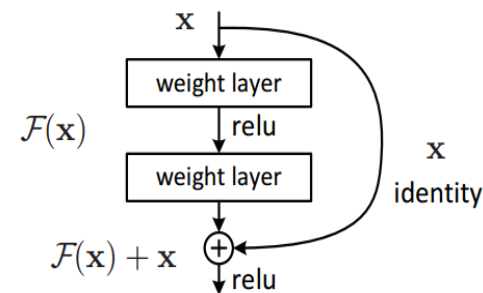
Experimental Results -model-

- DenseNet [Zhuang+, 2016]
- Densely connection of network
- Ultra deep CNNs -100 layers or more-



CIFAR-10 3.74%

- **ResNet** [He+, 2015]
- Shortcut connection
- Ultra deep CNNs -100 layers or more-



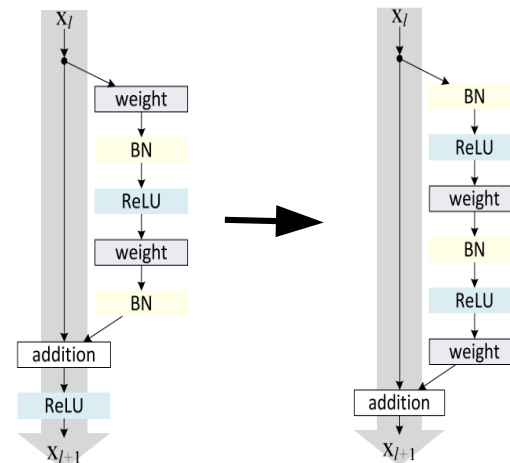
Residual block

+

CIFAR-10 6.61%

- **PreAct-ResNet** [He+, 2016]
- Improved version of ResNet

CIFAR-10 4.62%



Experimental Results **-dataset-**

► Datasets

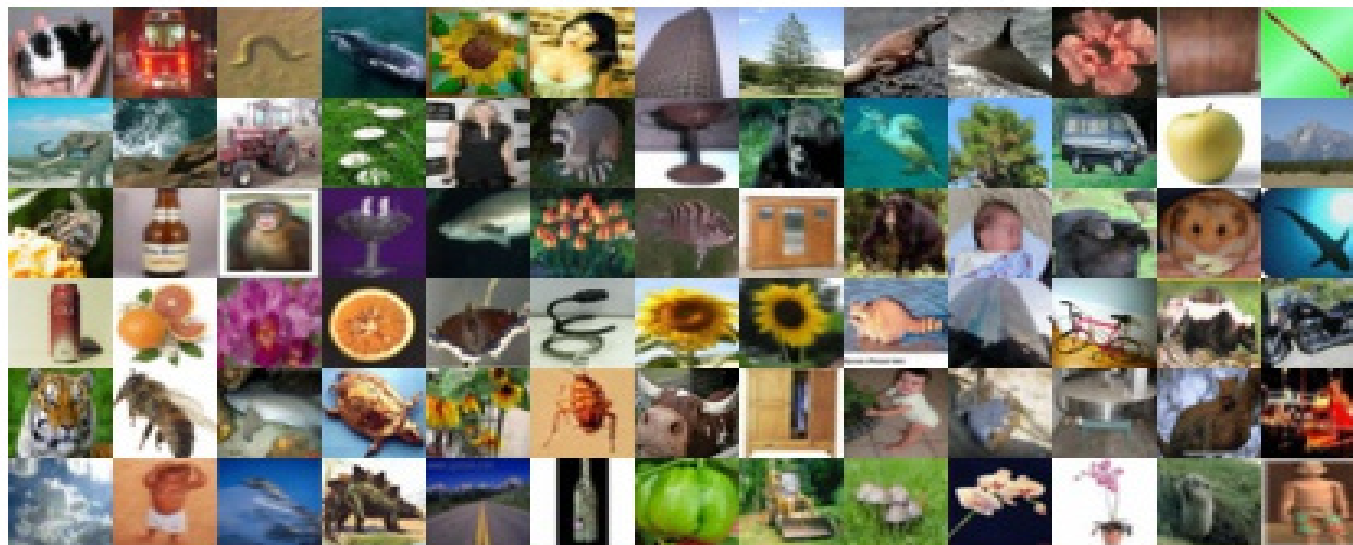
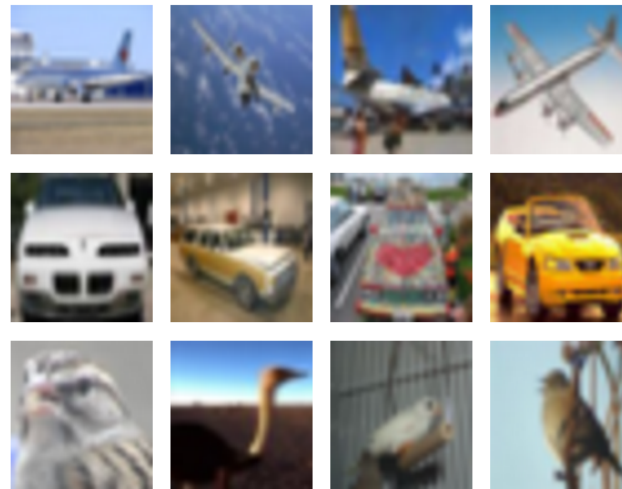
- **CIFAR-10** / **CIFAR-100**

10/**100** classes color images

image size : 32 x 32 x 3

training data : 50,000 images

test data : 10,000 images



Experimental Results **-dataset-**

► Datasets

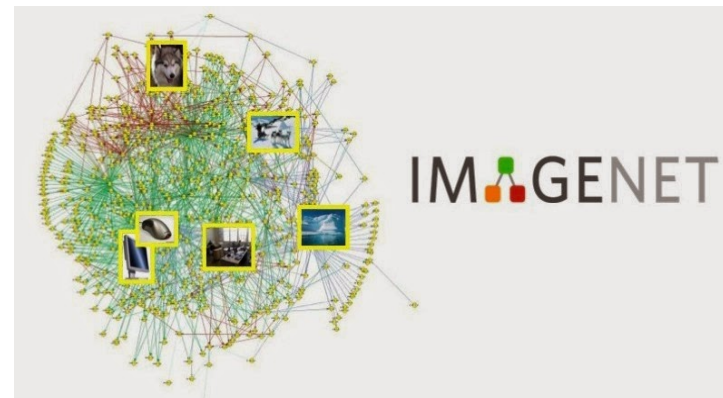
▪ **ImageNet**

1000 classes color images

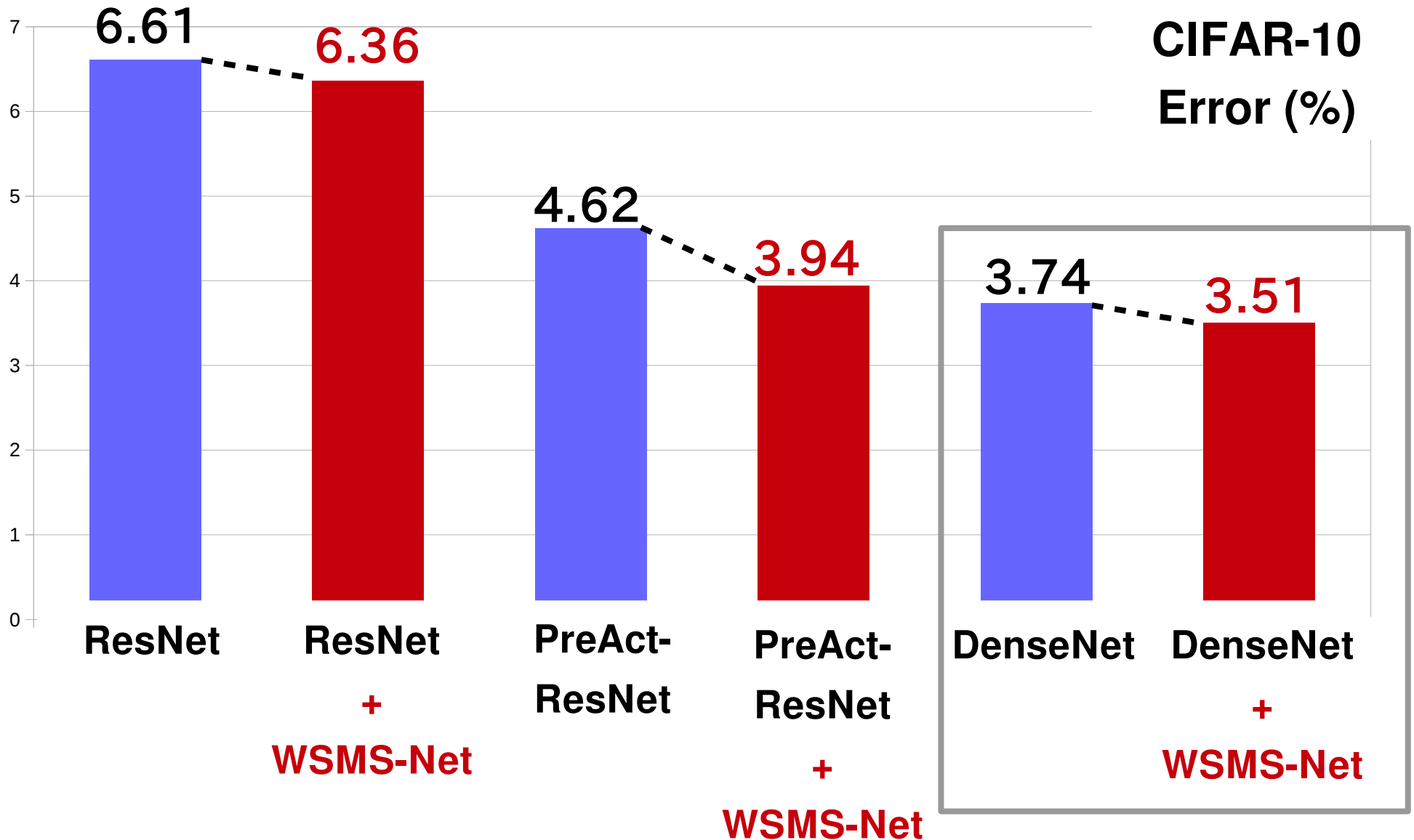
image size : **224 x 224 x 3**

training data : **1.28 million** images

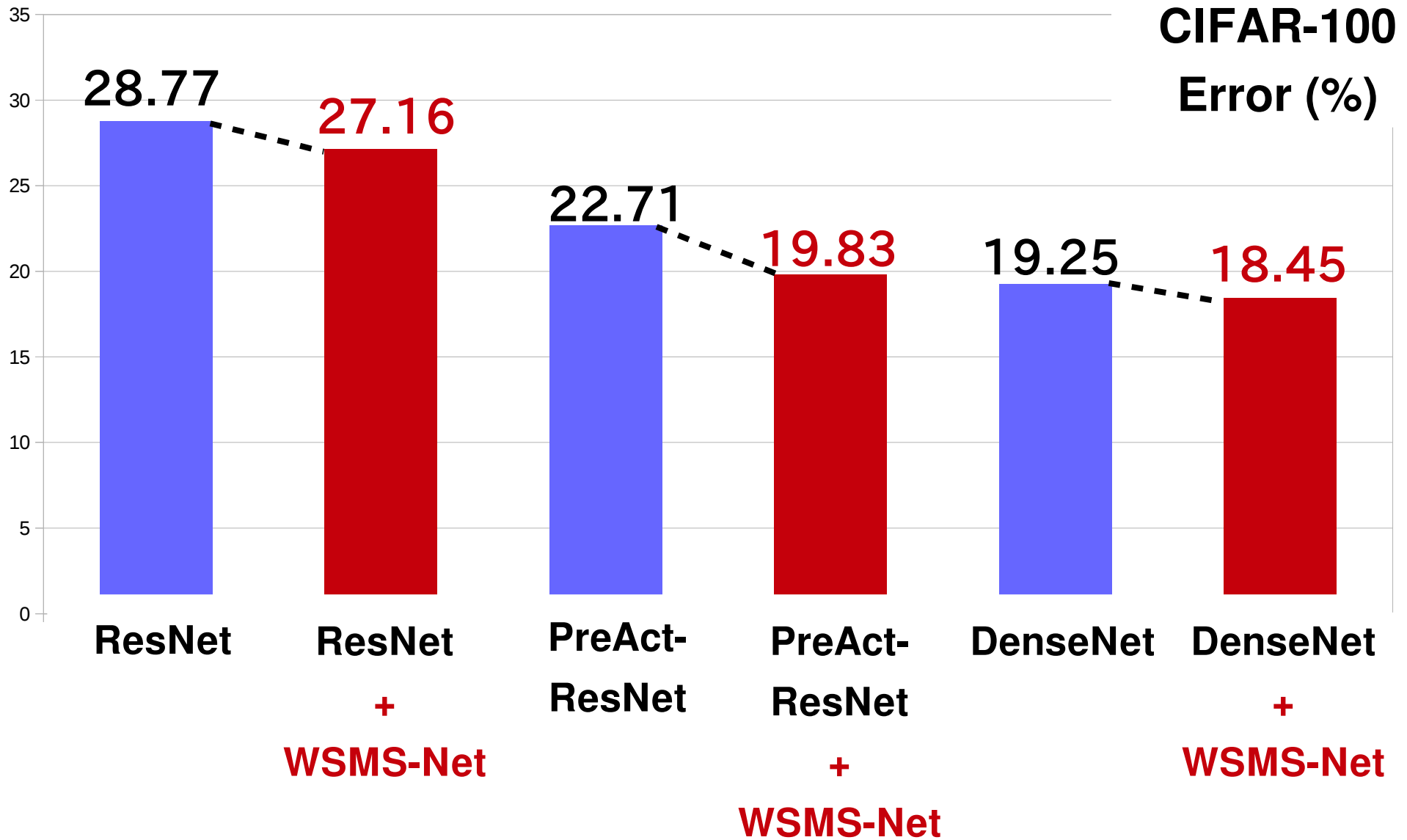
test data : **50,000** images



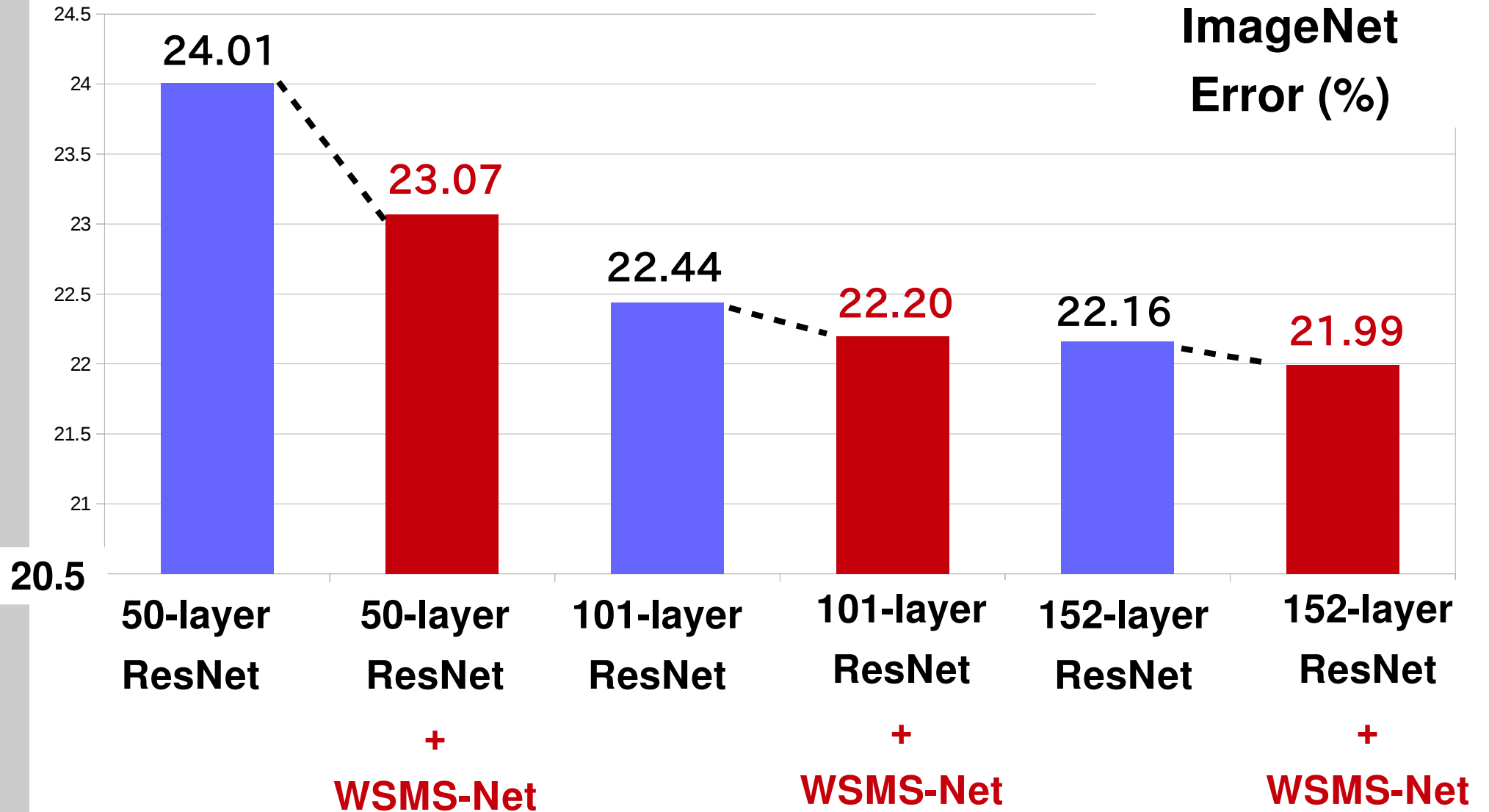
Experimental Results generalization capability



Experimental Results generalization capability



Experimental Results generalization capability



Index

▶ Introduction

▶ Our Method

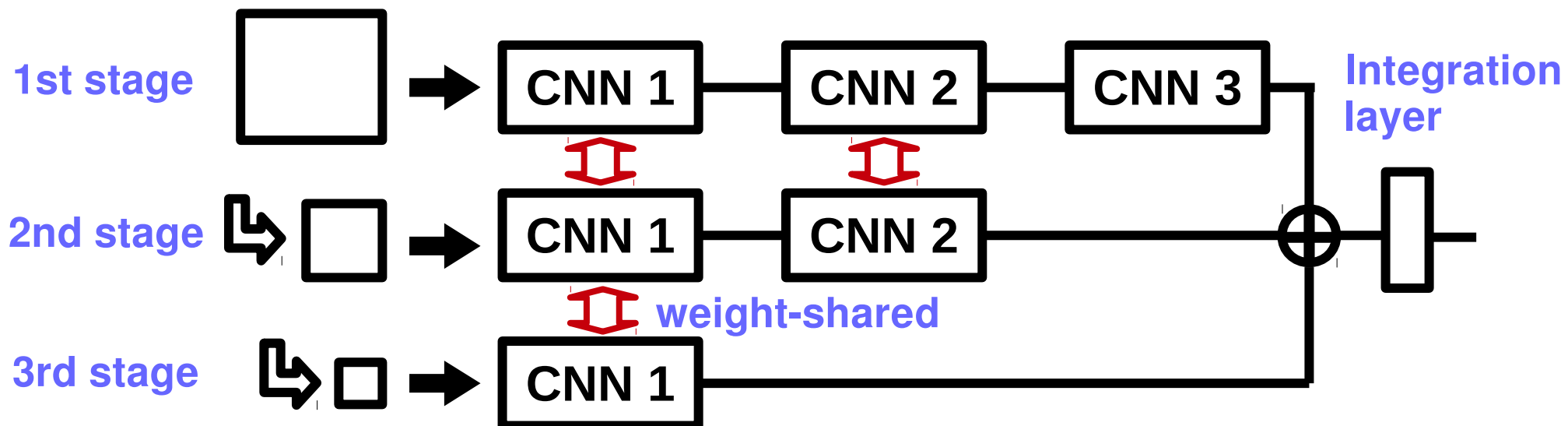
▶ **Experimental Results (new)**

- generalization capability (model, dataset)
- **number of stages**
- weight sharing

Experimental Results number of stages

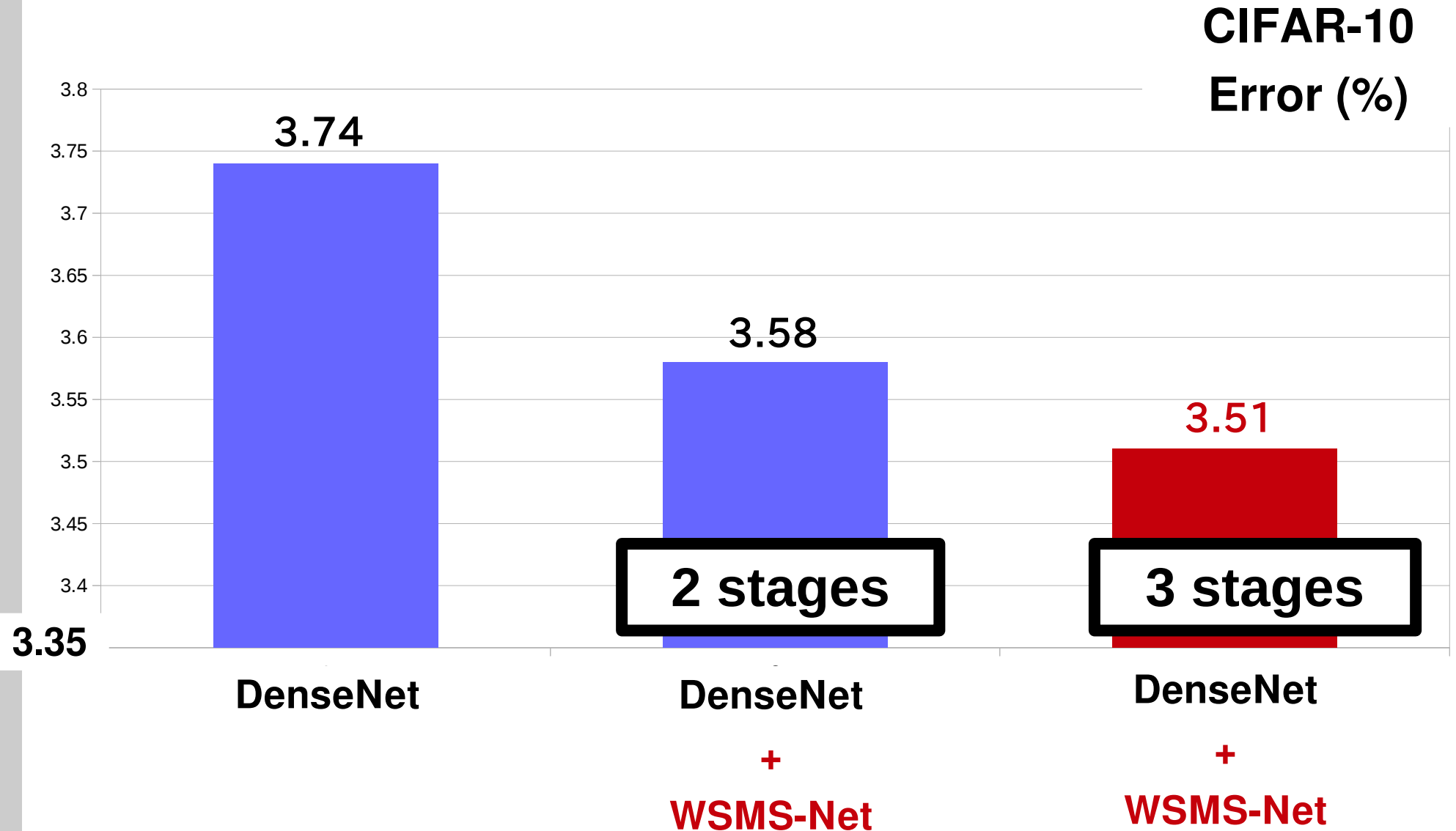
► Optimization of **number of stages**

- 2 stages on the CIFAR
- 4 stages on the ImageNet



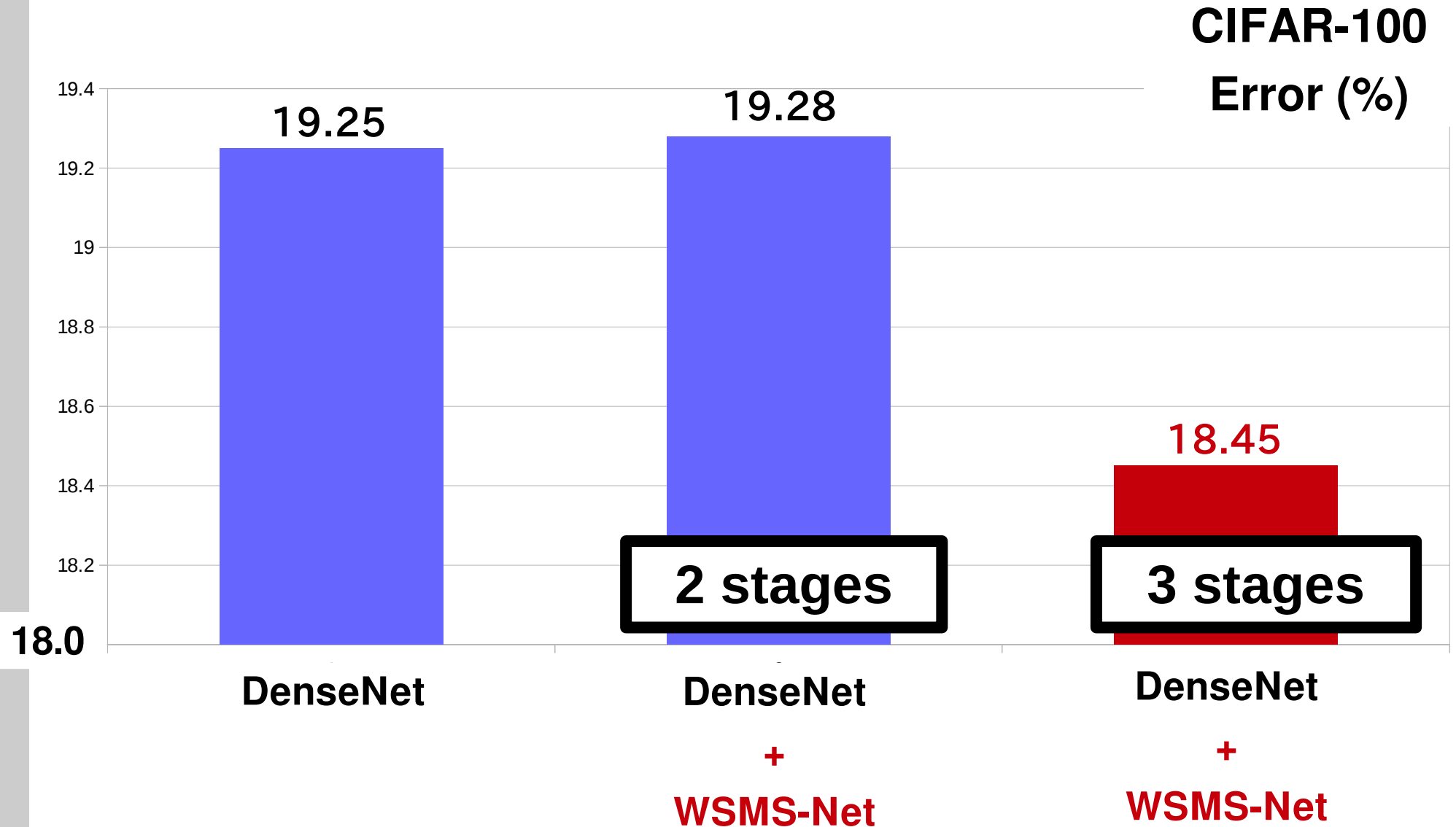
Experimental Results

number of stages



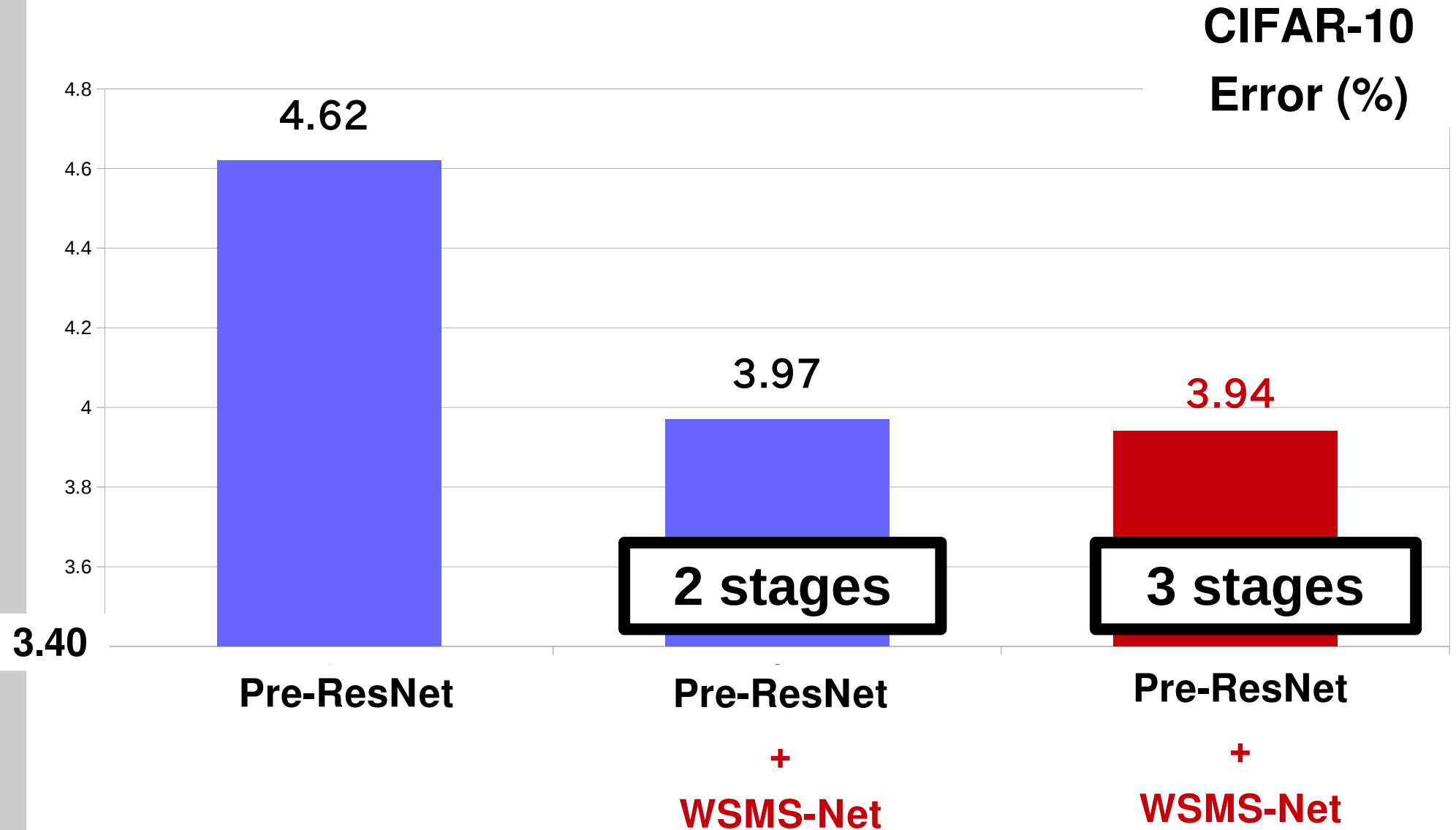
Experimental Results

number of stages



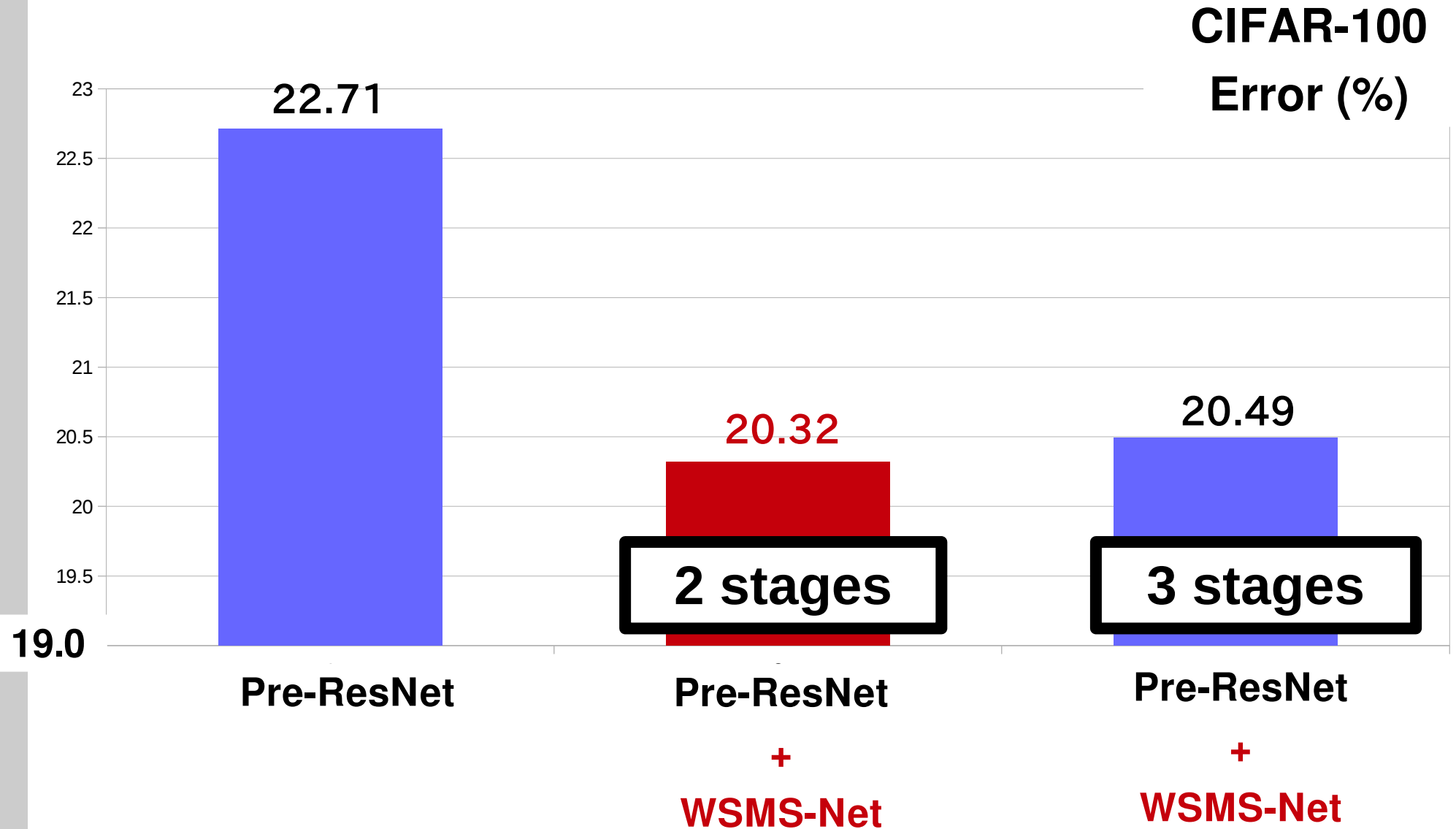
Experimental Results

number of stages

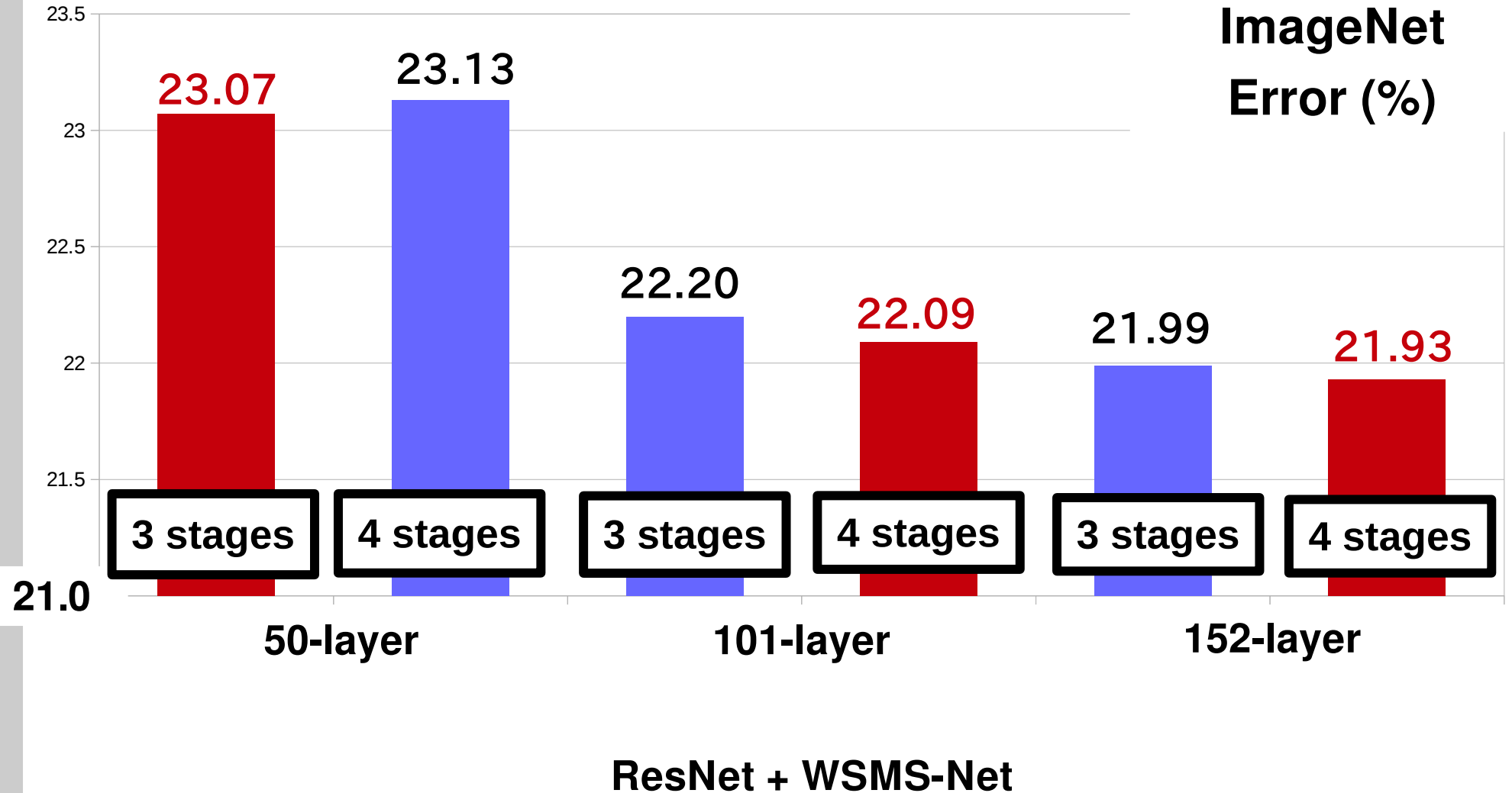


Experimental Results

number of stages



Experimental Results number of stages



Index

▶ Introduction

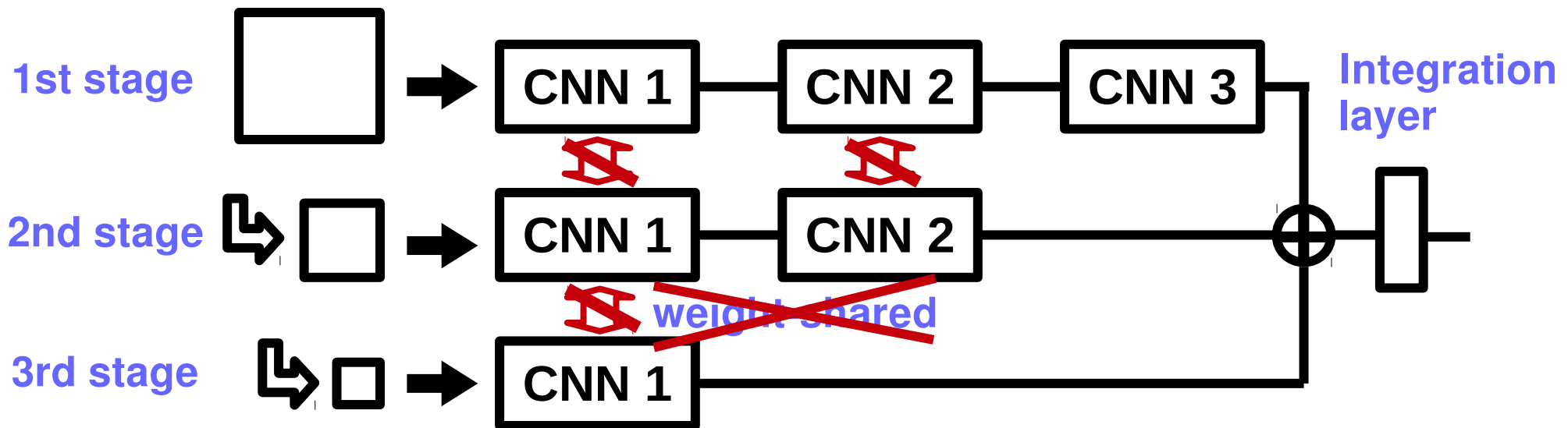
▶ Our Method

▶ **Experimental Results (new)**

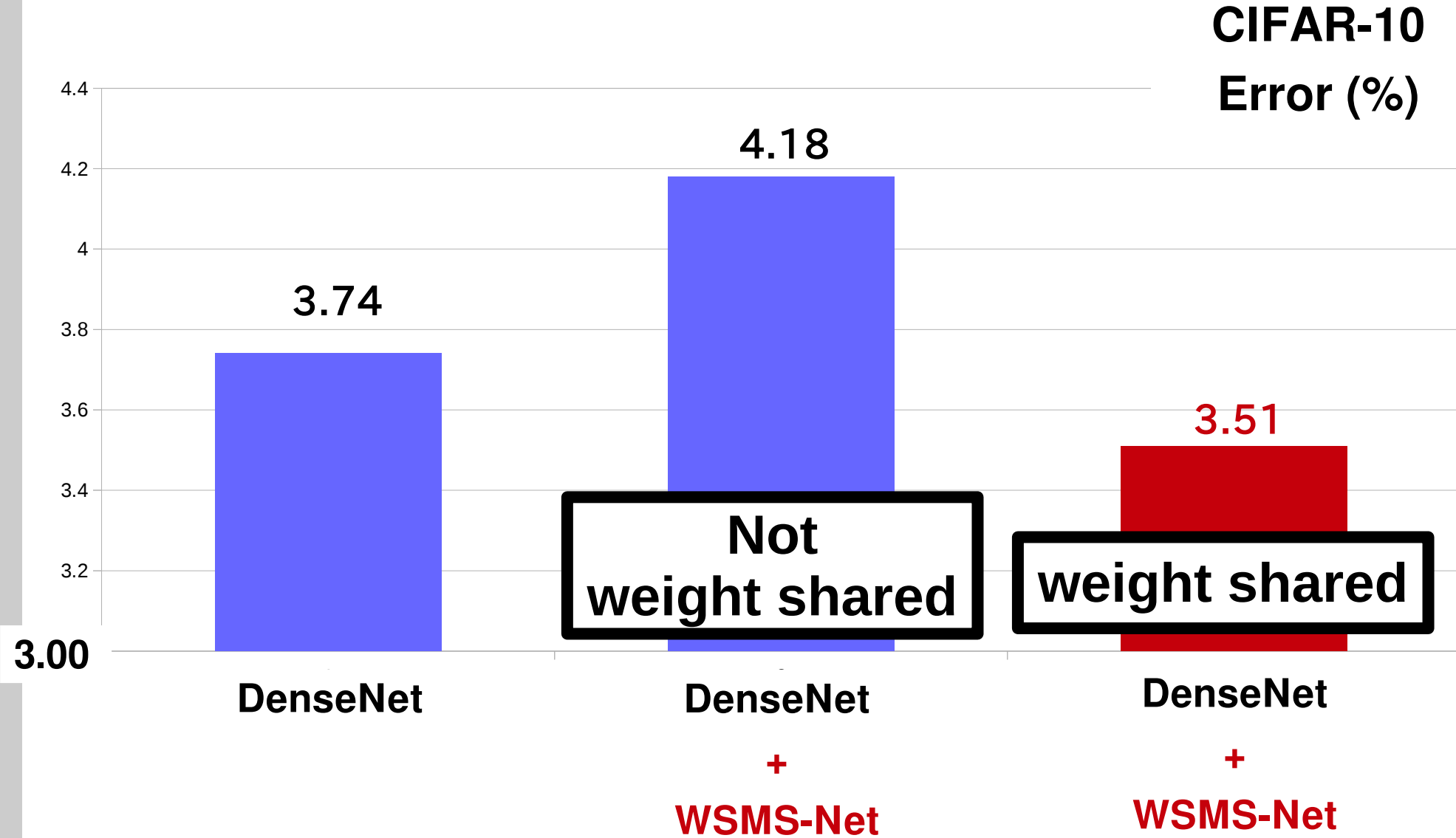
- generalization capability (model, dataset)
- number of stages
- **weight sharing**

Experimental Results weight sharing

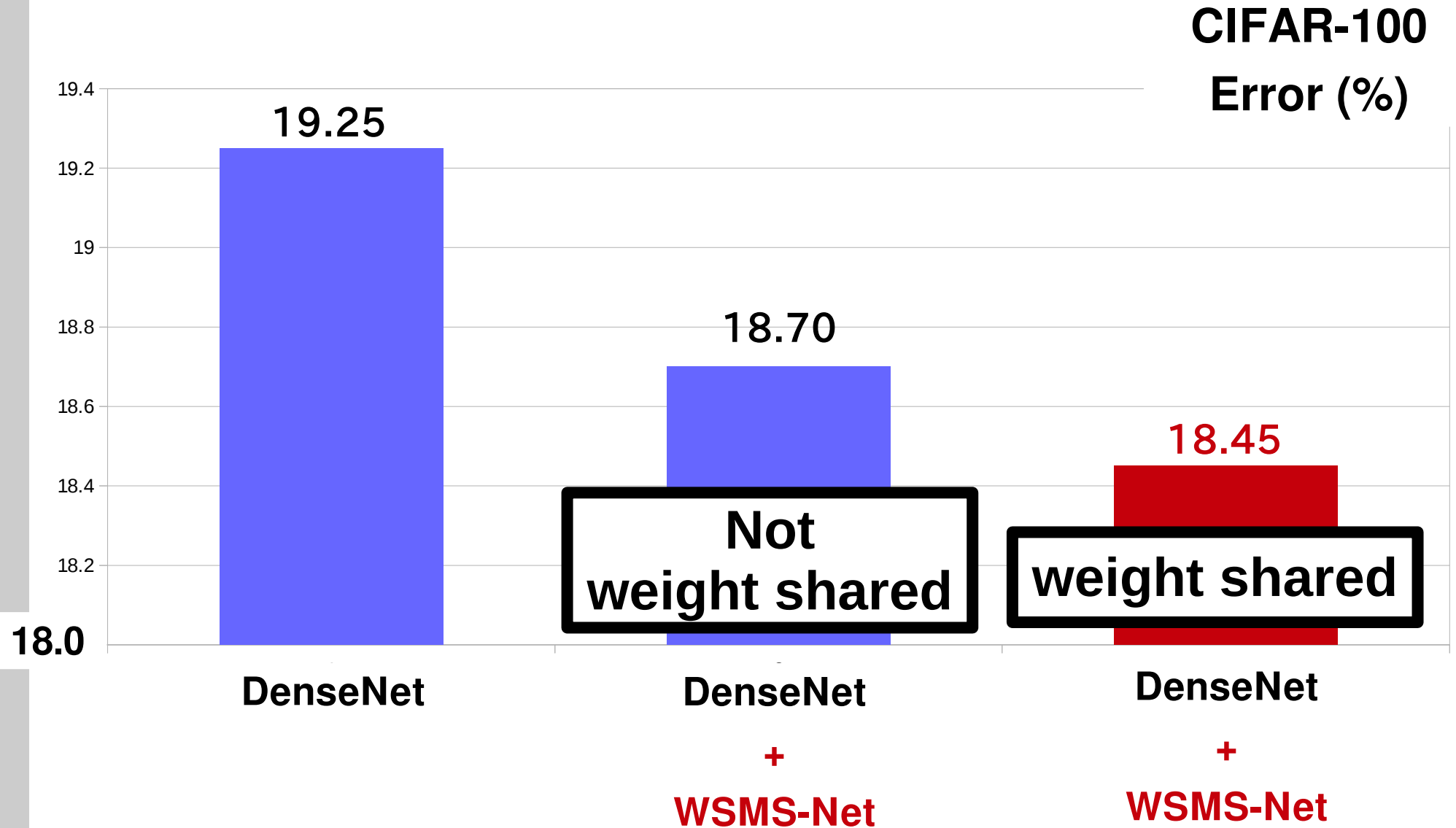
- ▶ The weight sharing really works? / is necessary?
 - evaluate by not weight shared network (MS-Net)



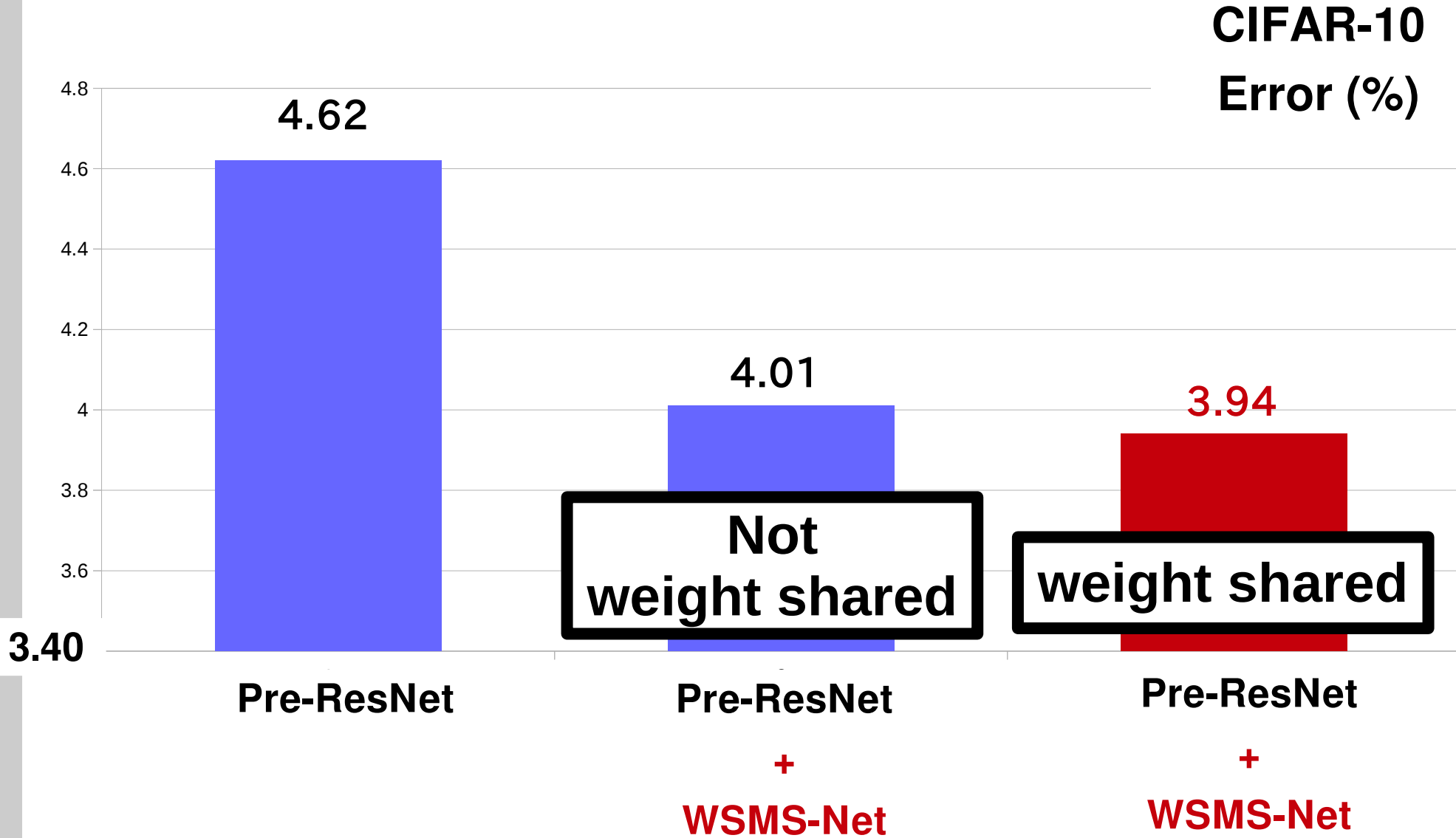
Experimental Results weight sharing



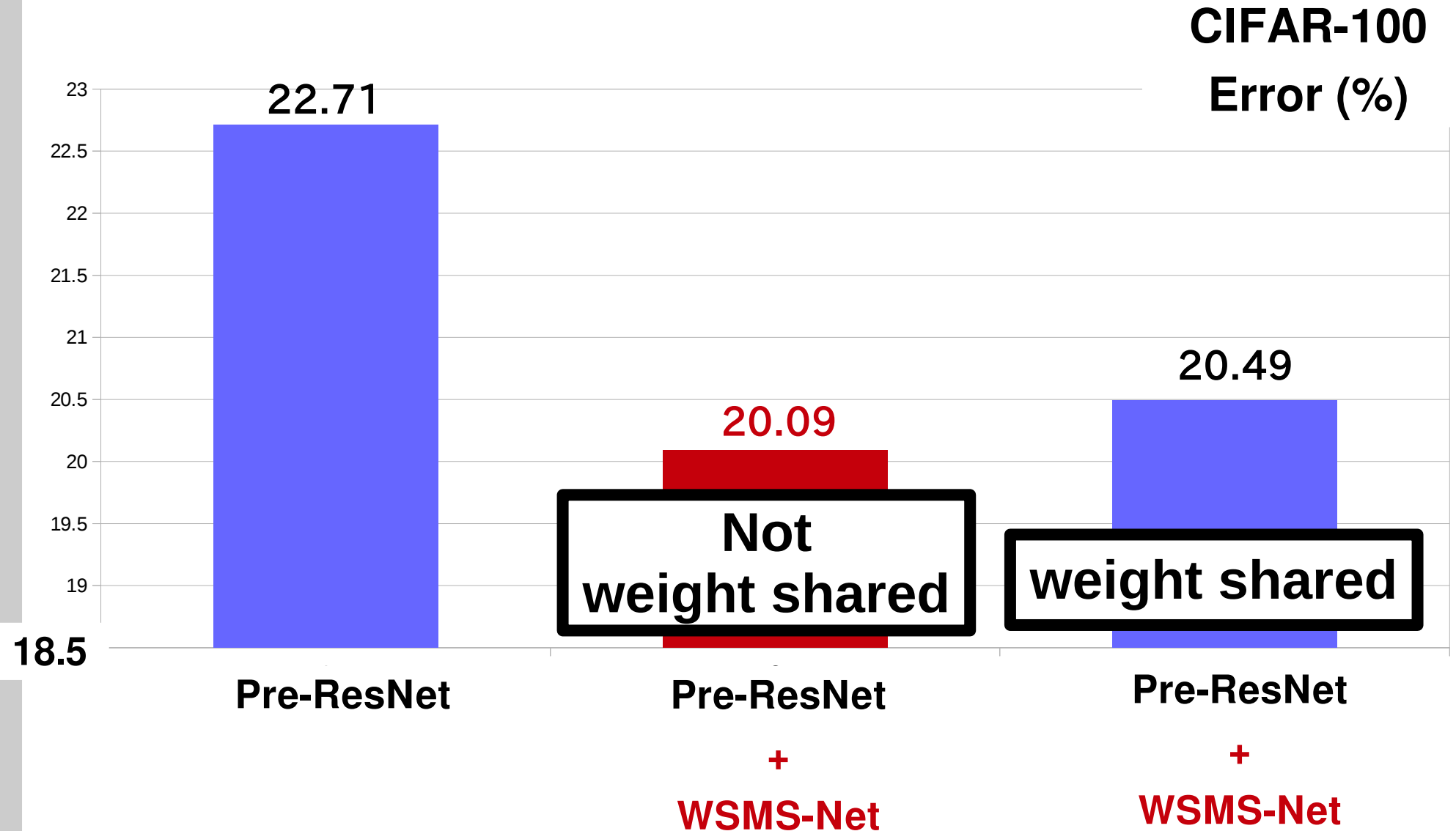
Experimental Results weight sharing



Experimental Results weight sharing



Experimental Results weight sharing

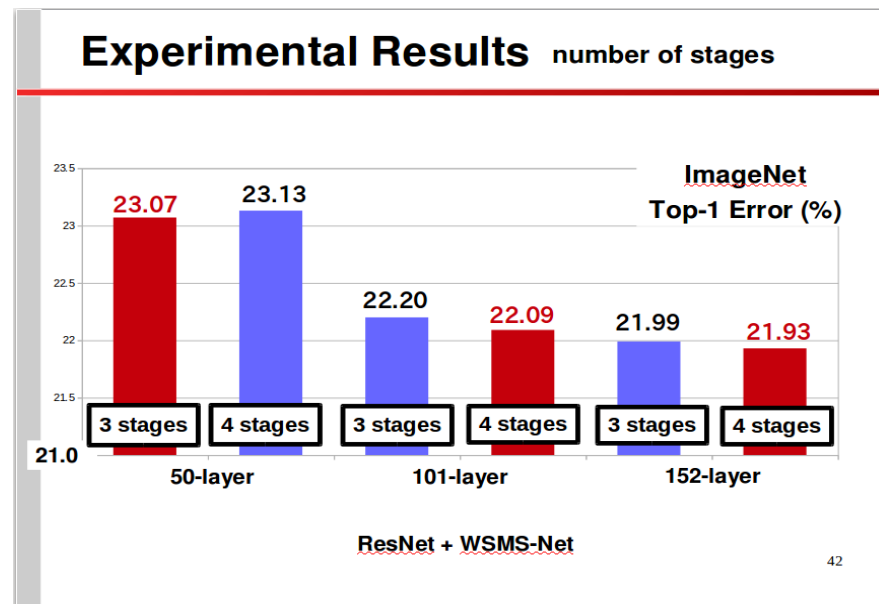
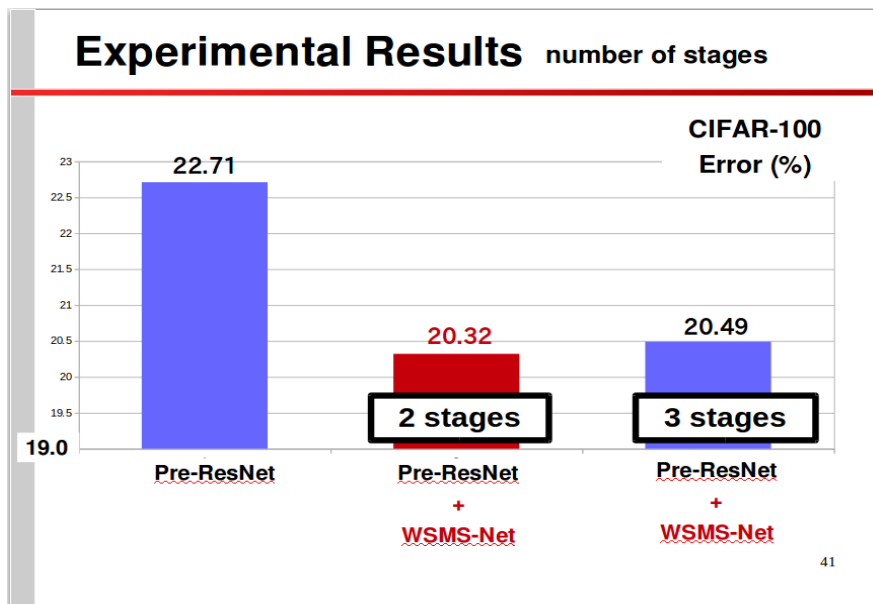
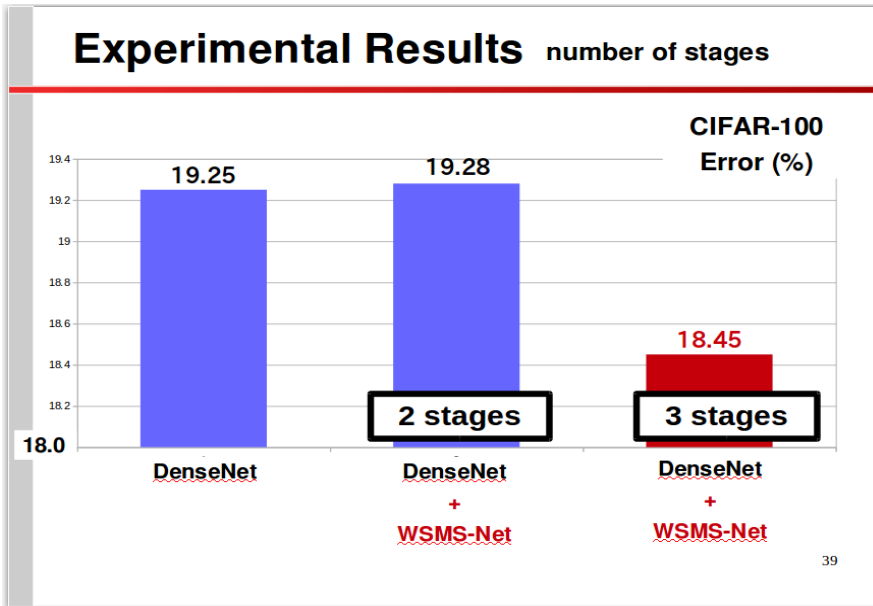


Index

- ▶ Introduction
- ▶ Our Method
- ▶ Experimental Results
- ▶ **Future Works**

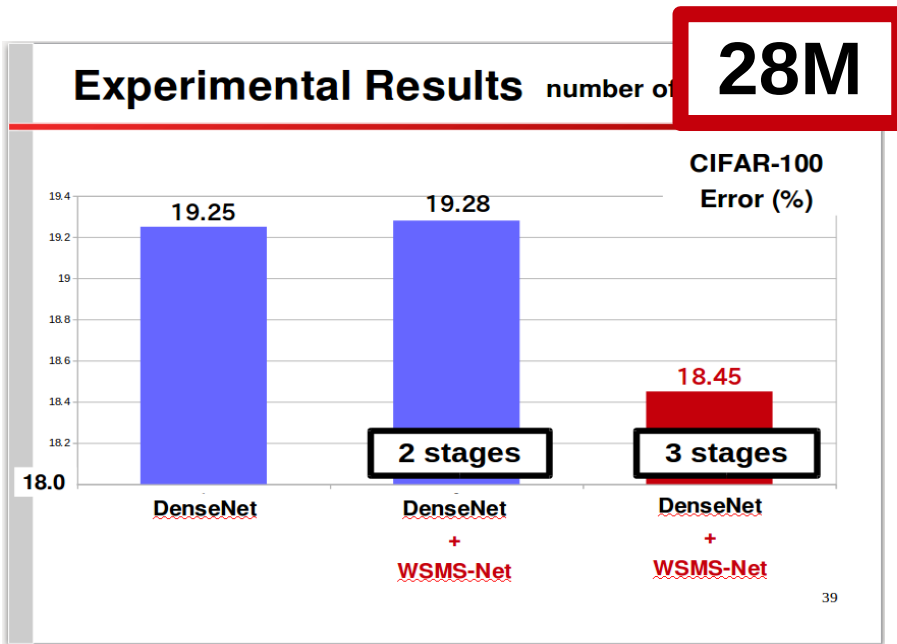
Future Works

number of stages



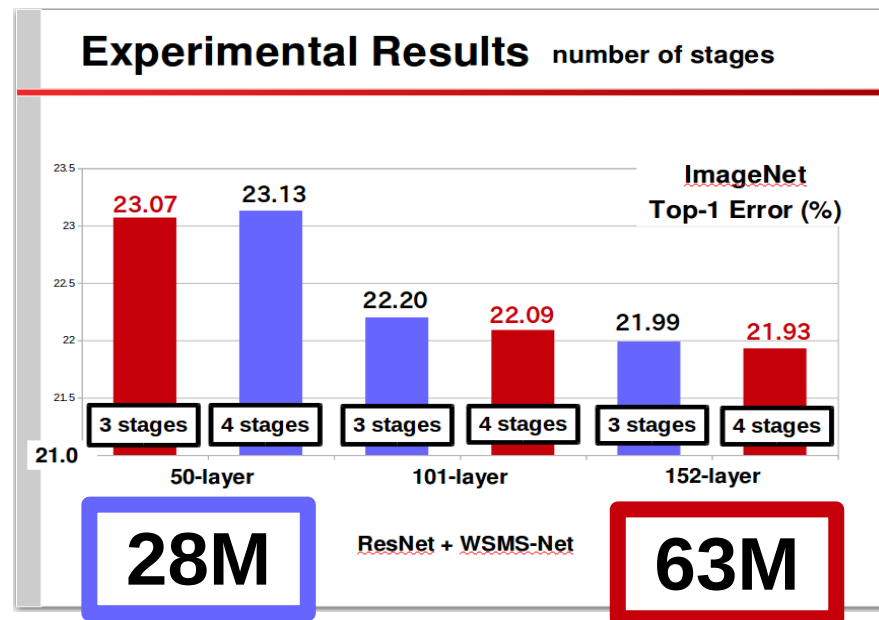
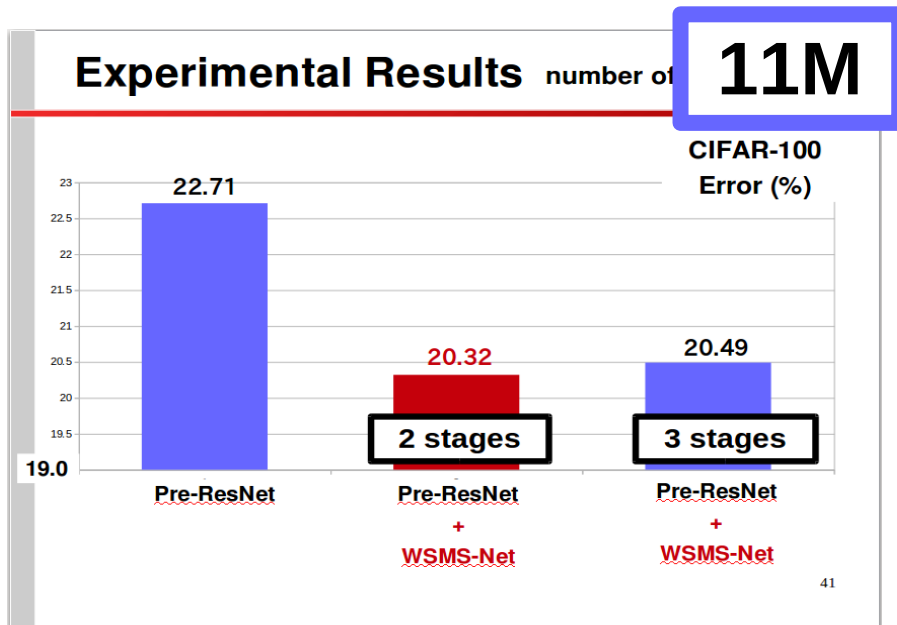
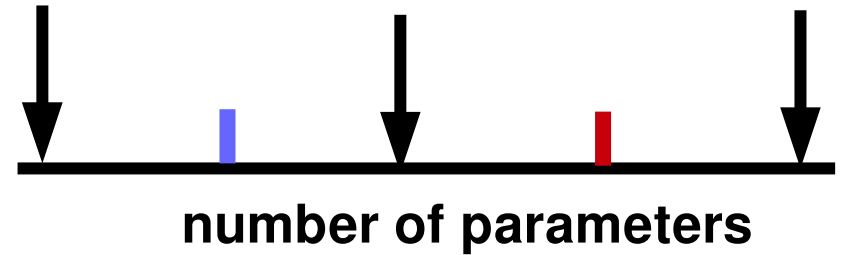
Future Works

number of stages



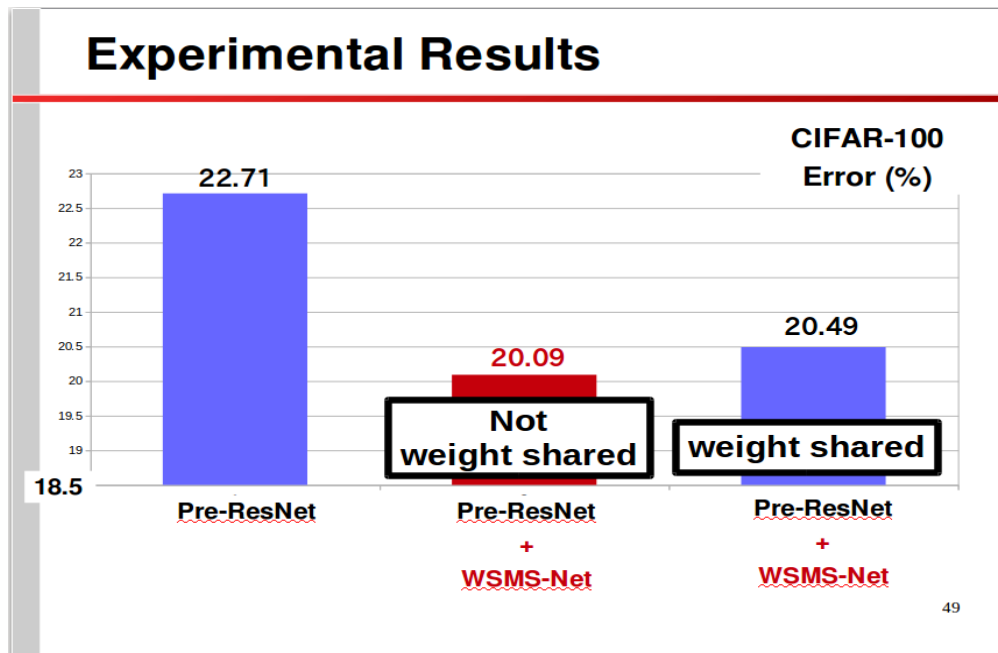
enough parameters : good effect

not enough parameters : bad effect



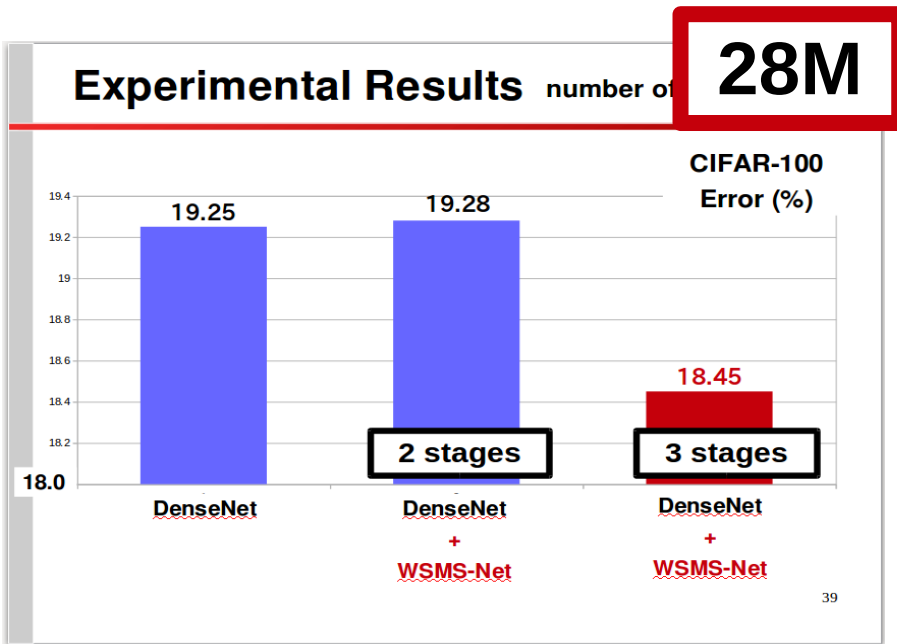
Future Works

weight sharing



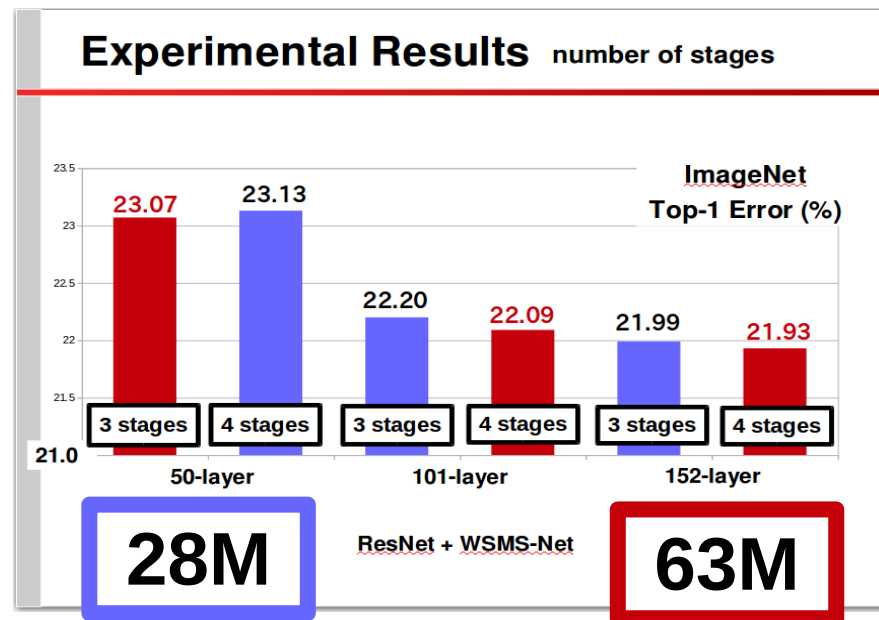
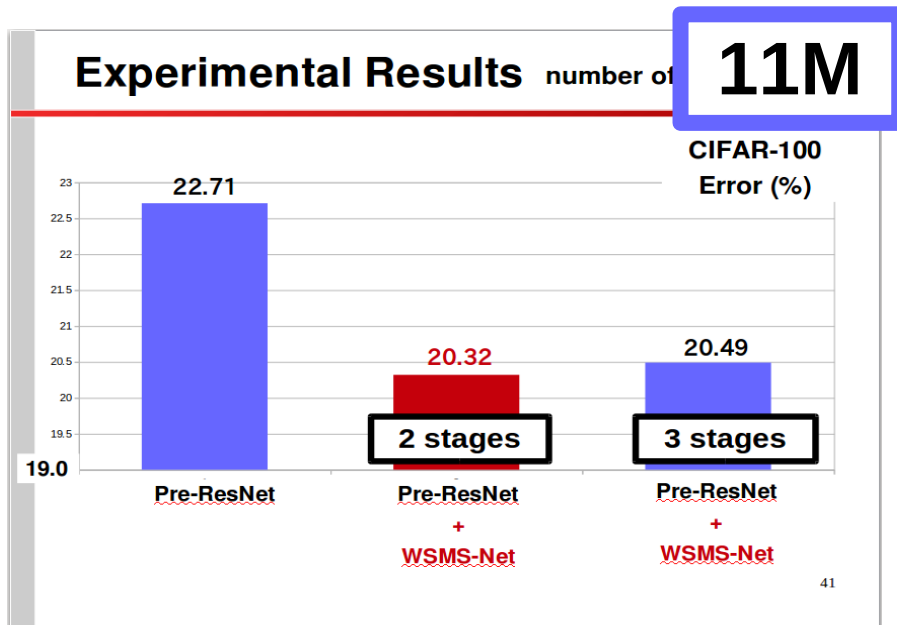
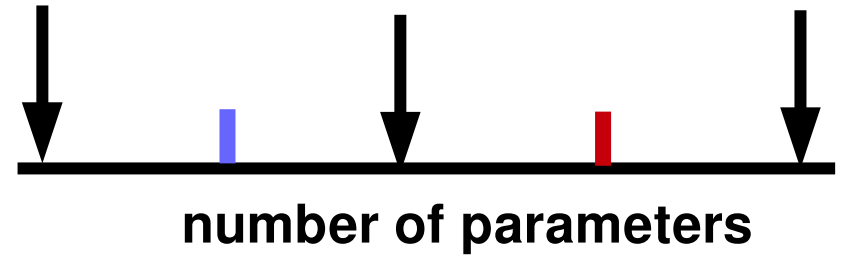
Future Works

number of stages



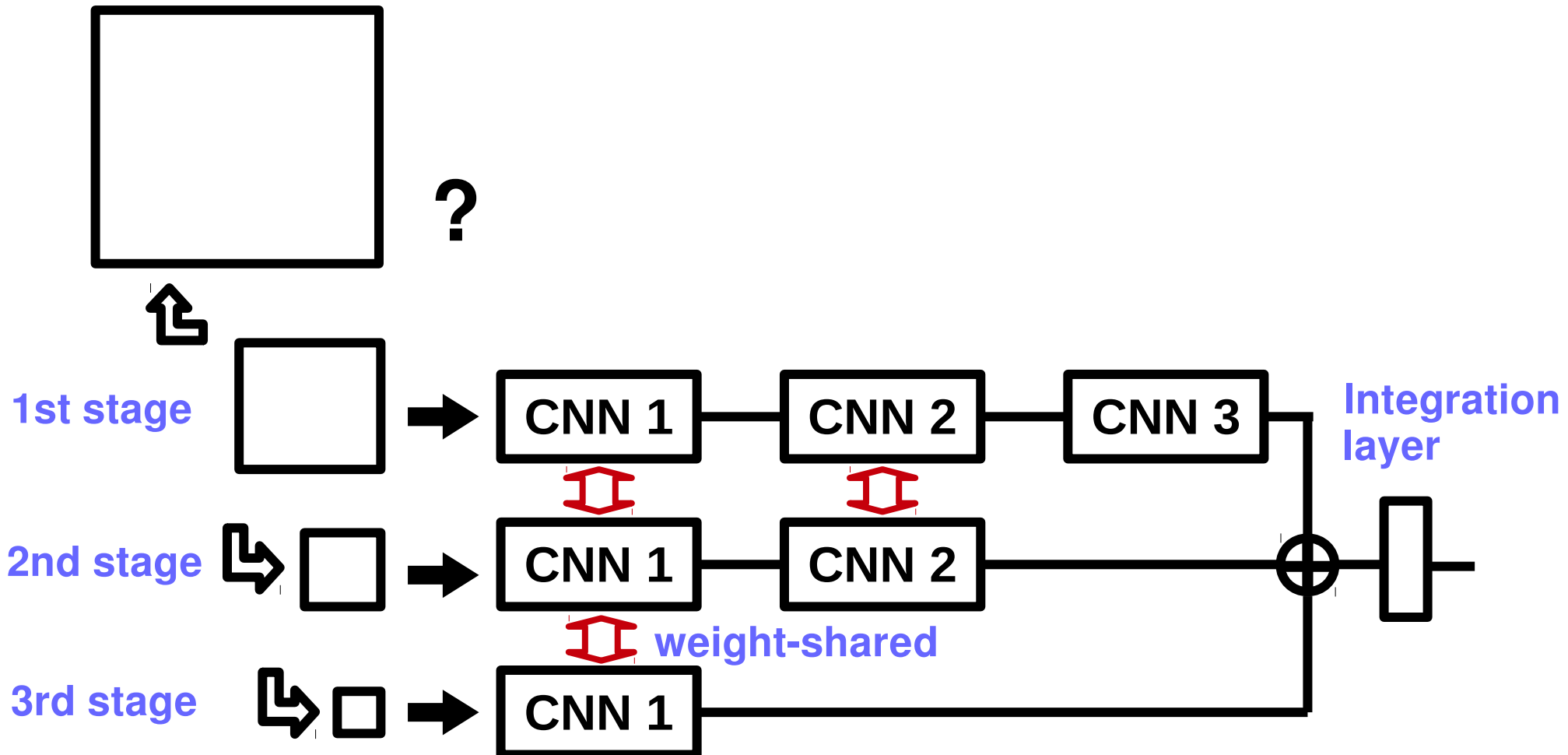
enough parameters : good effect

not enough parameters : bad effect



Future Works

► Model Extension



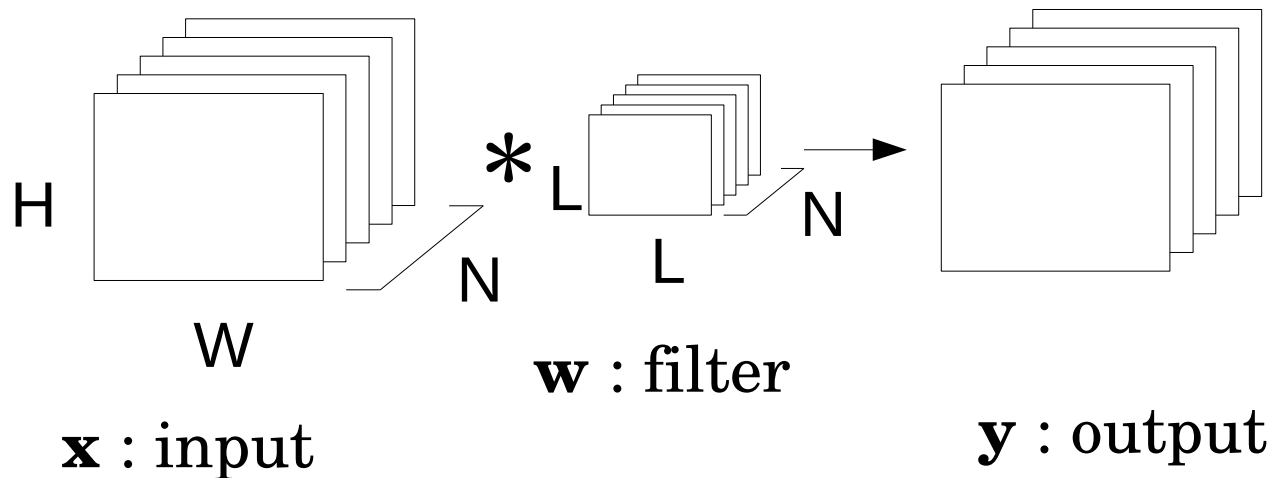
Appendix

Conclusion

- ▶ **WSMS-Net is scale-invariant, or at least robust to the scaling of object.**
- ▶ **WSMS-Net achieved the better performance than the Existing CNNs.**
- ▶ **Scale invariance is effective for better classification**

What is the CNN?

□ Convolution



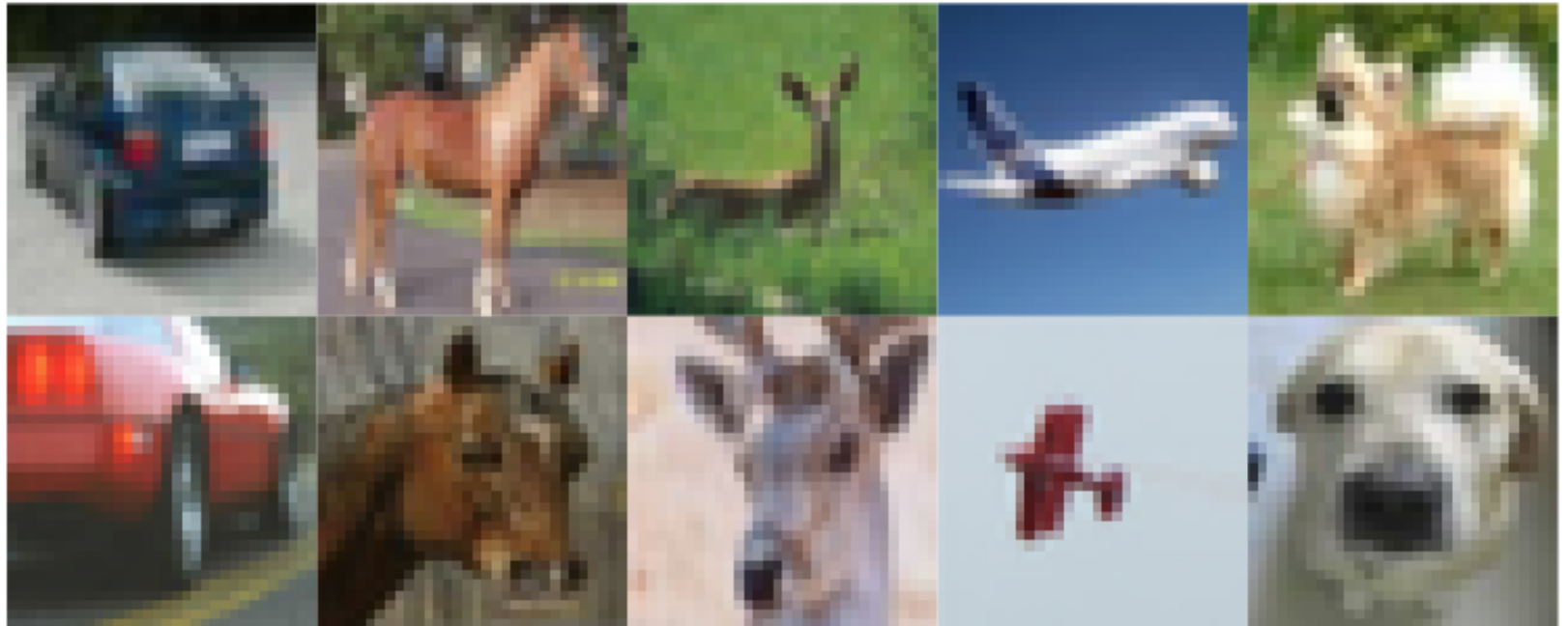
$$y_{ijk} = \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} w_{mnik} x_{(i+m)(j+n)k} + b_k$$

ijk : height, width, channel of image

Introduction

- Examples of scaled type images of CIFAR-10

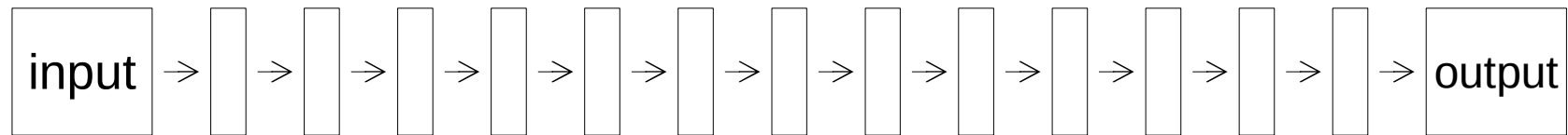
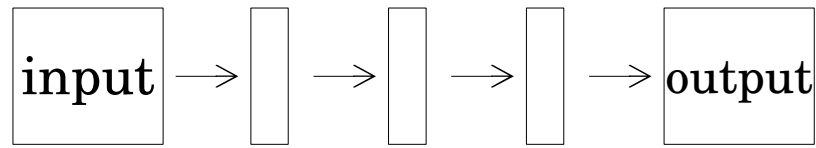
popular



scaled

Introduction

□ Gradient vanishing problem



Deepening the layers results in the unreached of gradient information

Related works

□ ResNet family CNNs we use

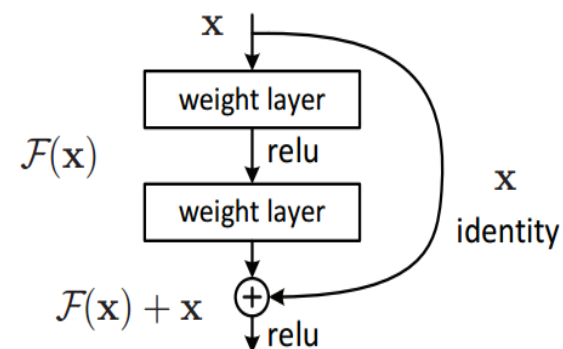
▶ ResNet [He+,2015]

- Shortcut connection
- Ultra deep CNNs -100 layers or more-

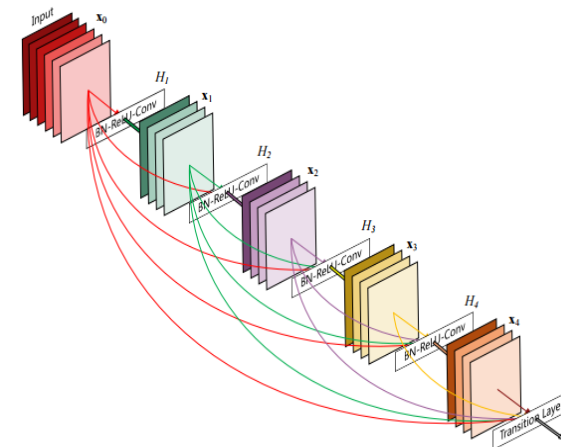
▶ DenseNet [Zhuang+,2016]

- Densely connection of network
- Ultra deep CNNs -100 layers or more-

overcome the gradient vanishing problem



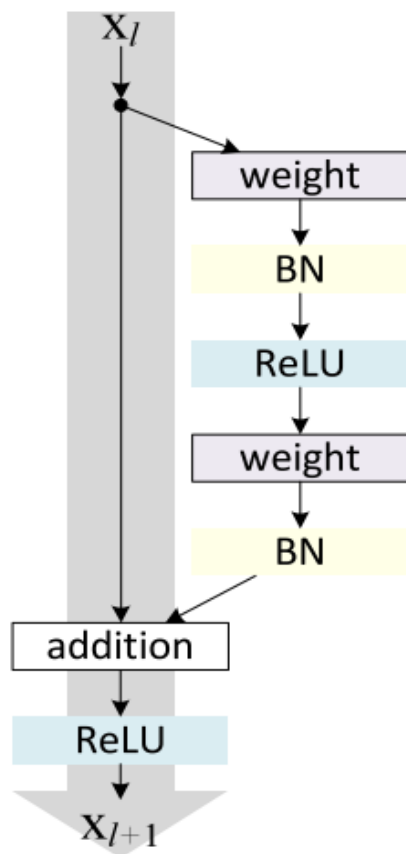
Residual block



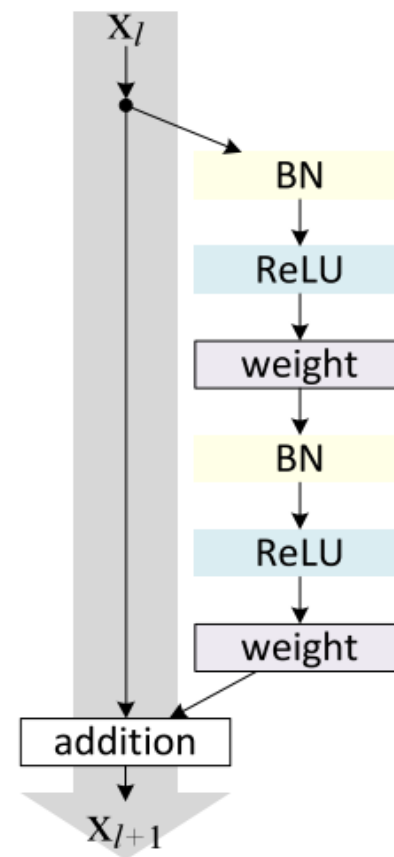
Related works

► PreAct-ResNet [He+,2016]

- Different order of convolution, BatchNorm and ReLU



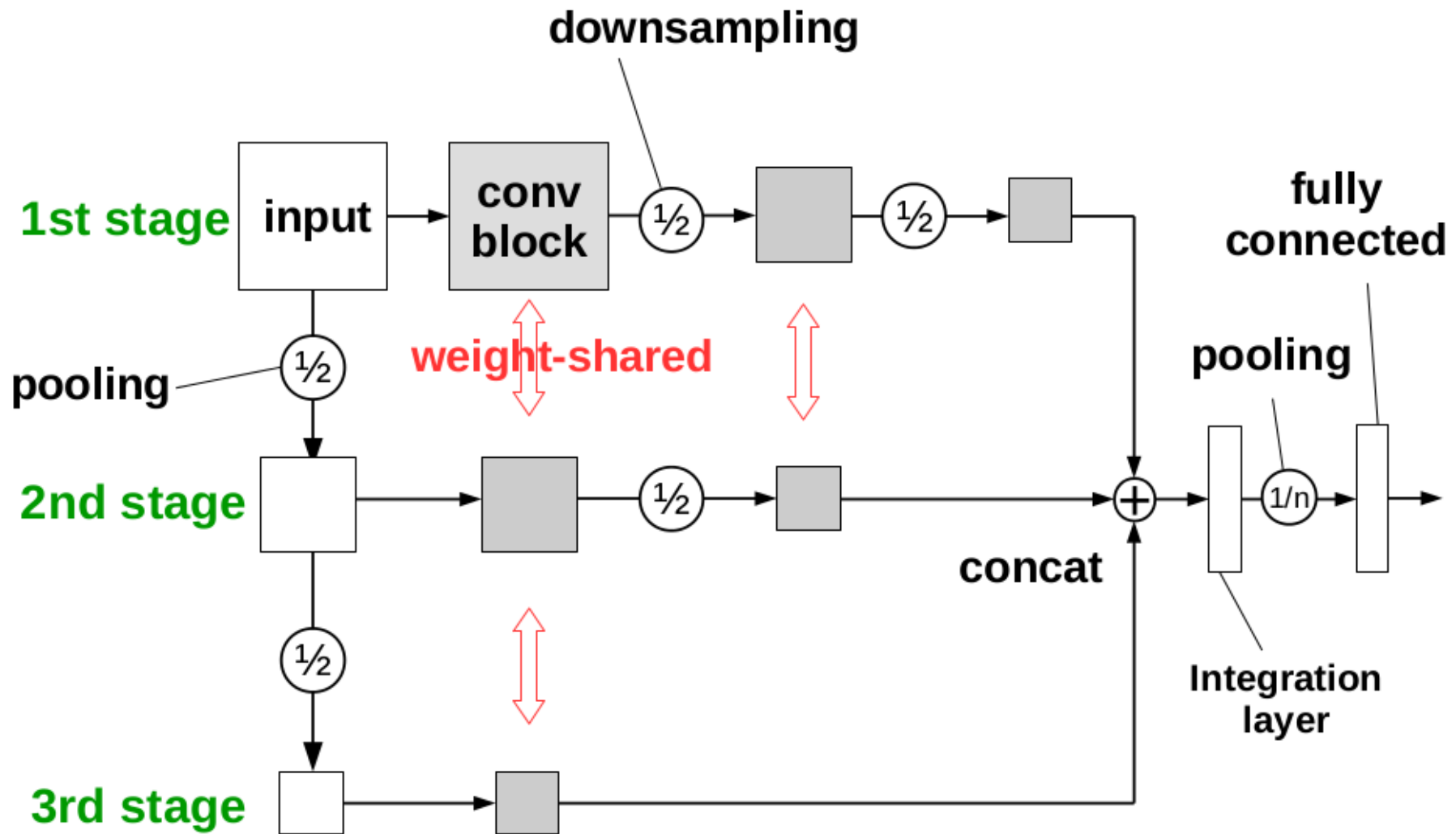
original



PreAct-ResNet

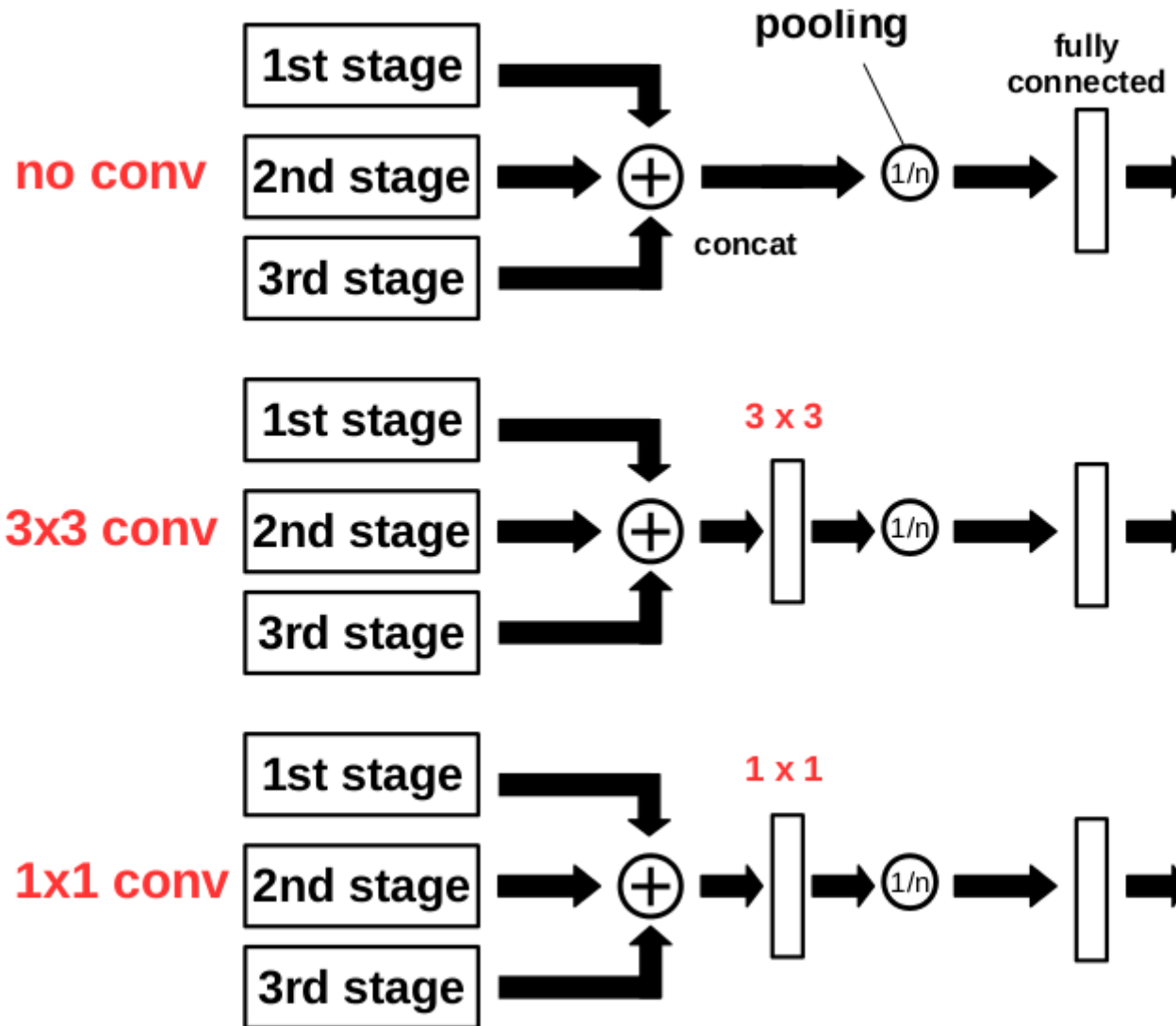
Proposed method -network-

► Basic architecture of WSMS-Net with 3 stages



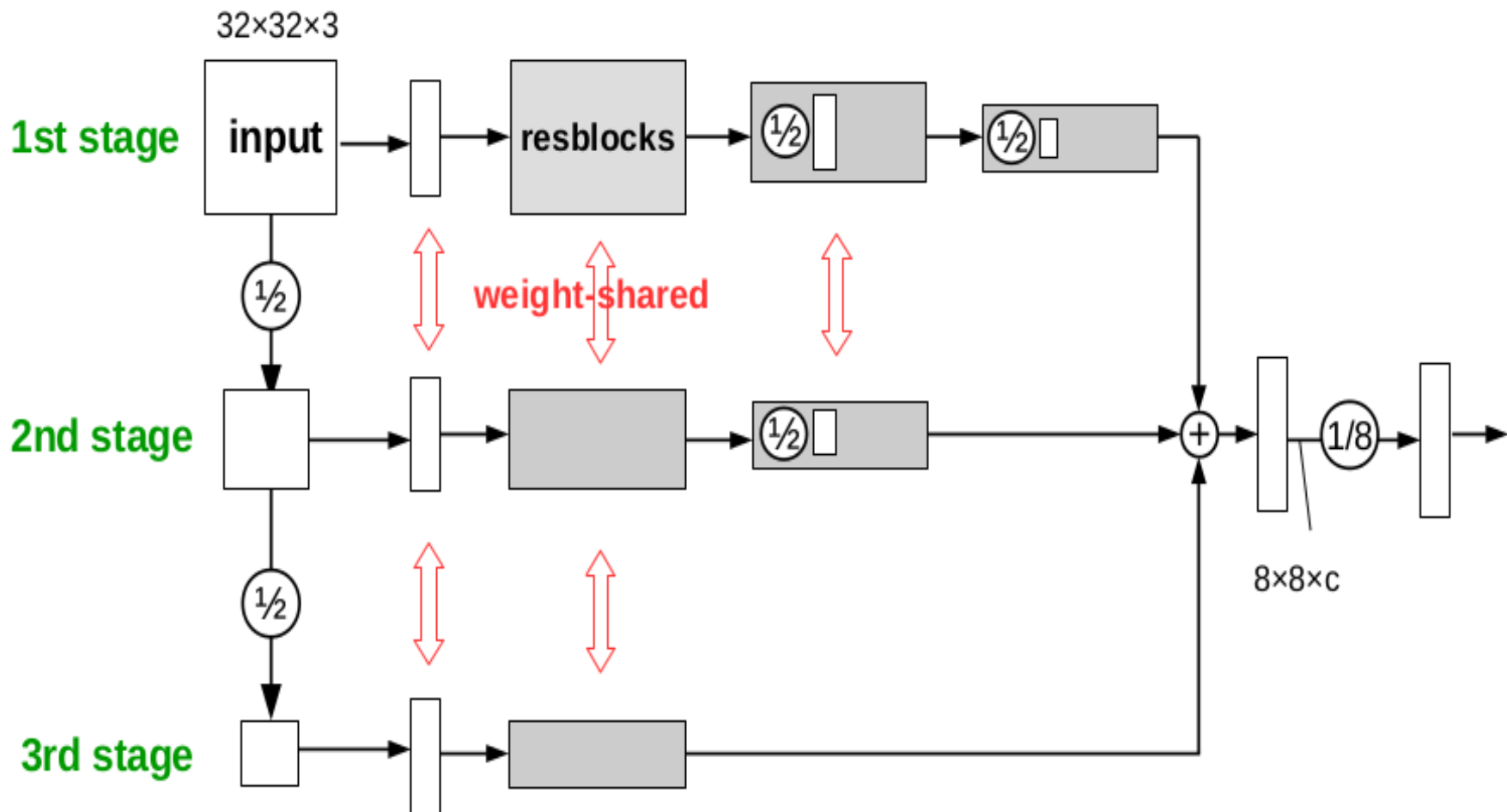
Proposed method -integration layer-

- We prepare 3 types of integration layers



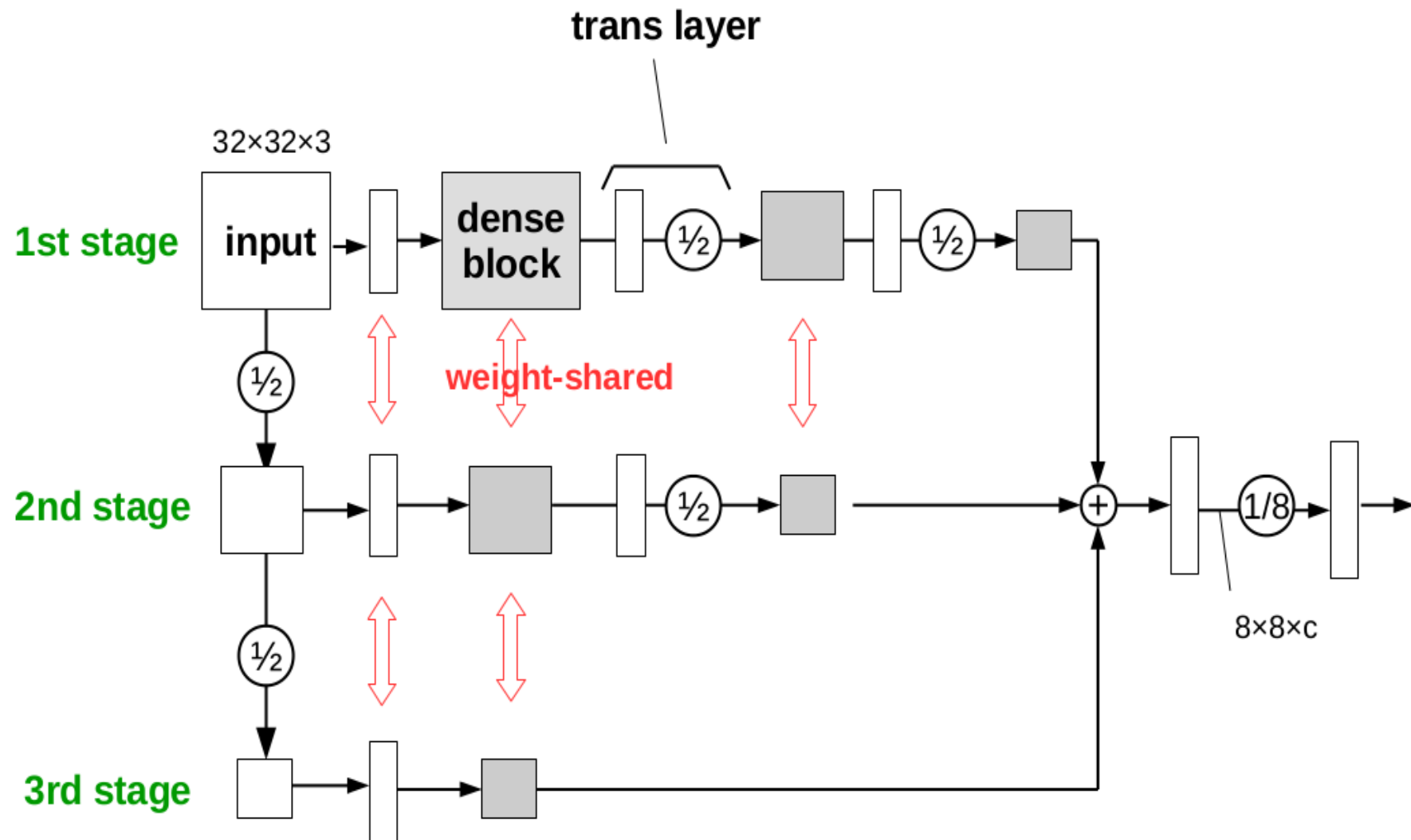
Experimental Results -network-

► WSMS-Net with ResNet



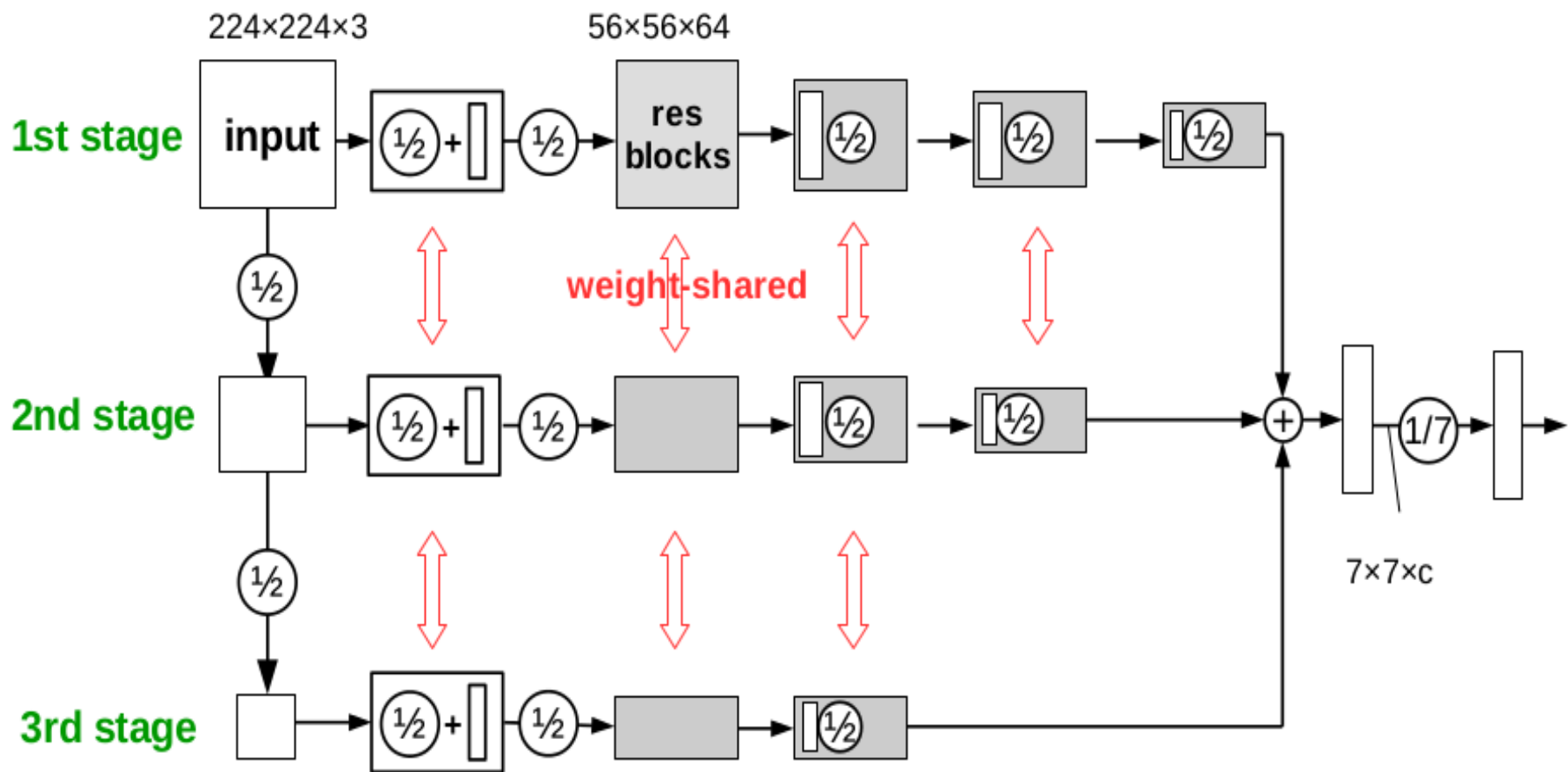
Experimental Results -network-

► WSMS-Net with DenseNet



Experimental Results -network-

► WSMS-Net with ResNet for ImageNet



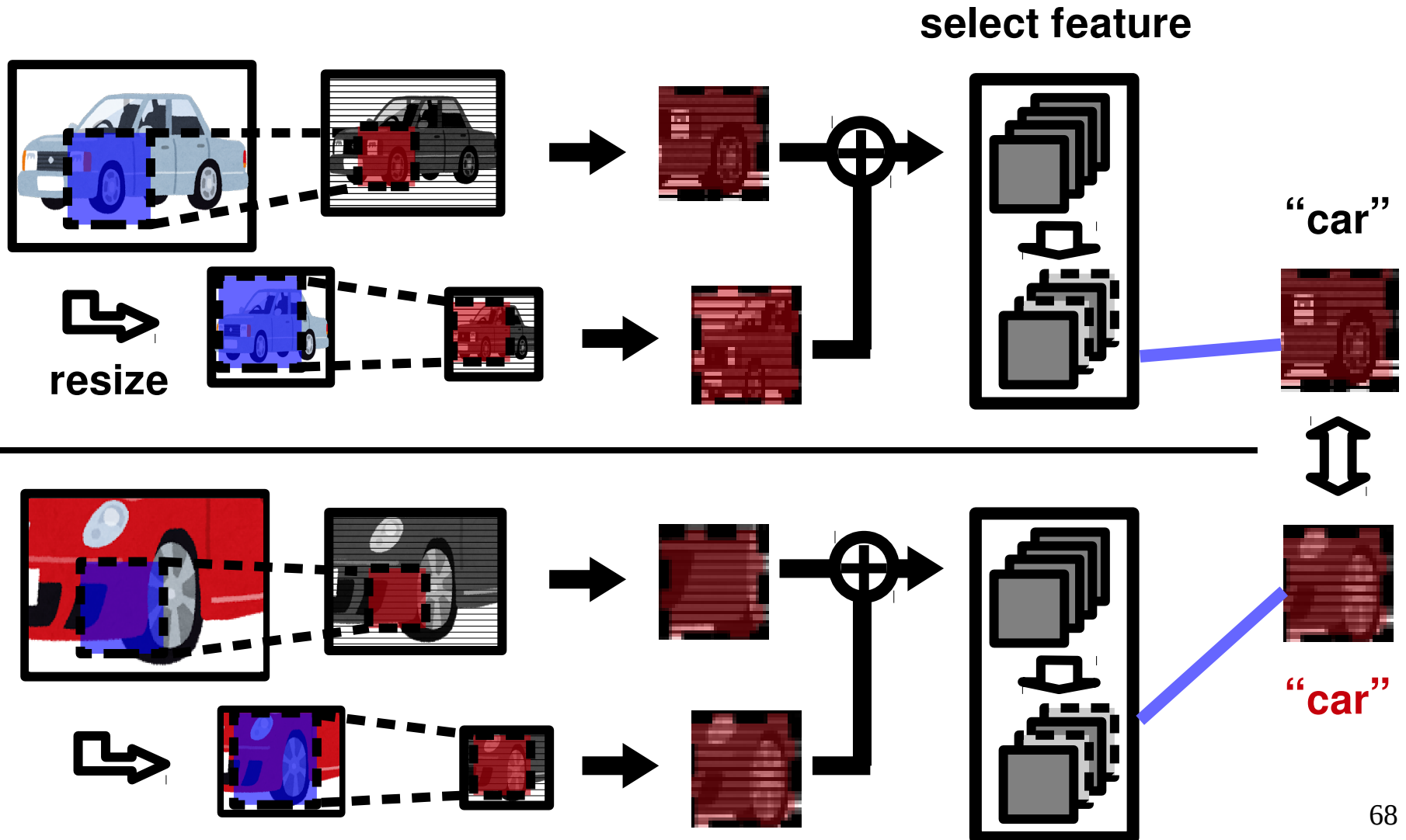
Experimental Results -classification accuracy-

- Detail results on the **ImageNet** with **ResNet + WSMS-Net**
- results for optimization of **the number of stages**

Network	depth	#params	top-1 Error(%)	top-5 Error(%)
ResNet	50	25.6M	24.01	7.02
WSMS-ResNet (3 stages, 1×1 conv)	51	28.3M	23.07	6.44
WSMS-ResNet (4 stages, 1×1 conv)	51	28.5M	23.13	6.57
ResNet	101	44.5M	22.44	6.21
WSMS-ResNet (3 stages, 1×1 conv)	102	47.3M	22.20	6.22
WSMS-ResNet (4 stages, 1×1 conv)	102	47.6M	22.09	6.06
ResNet	152	60.2M	22.16	6.16
WSMS-ResNet (3 stages, 1×1 conv)	153	63.0M	21.99	6.04
WSMS-ResNet (4 stages, 1×1 conv)	153	63.3M	21.93	5.90

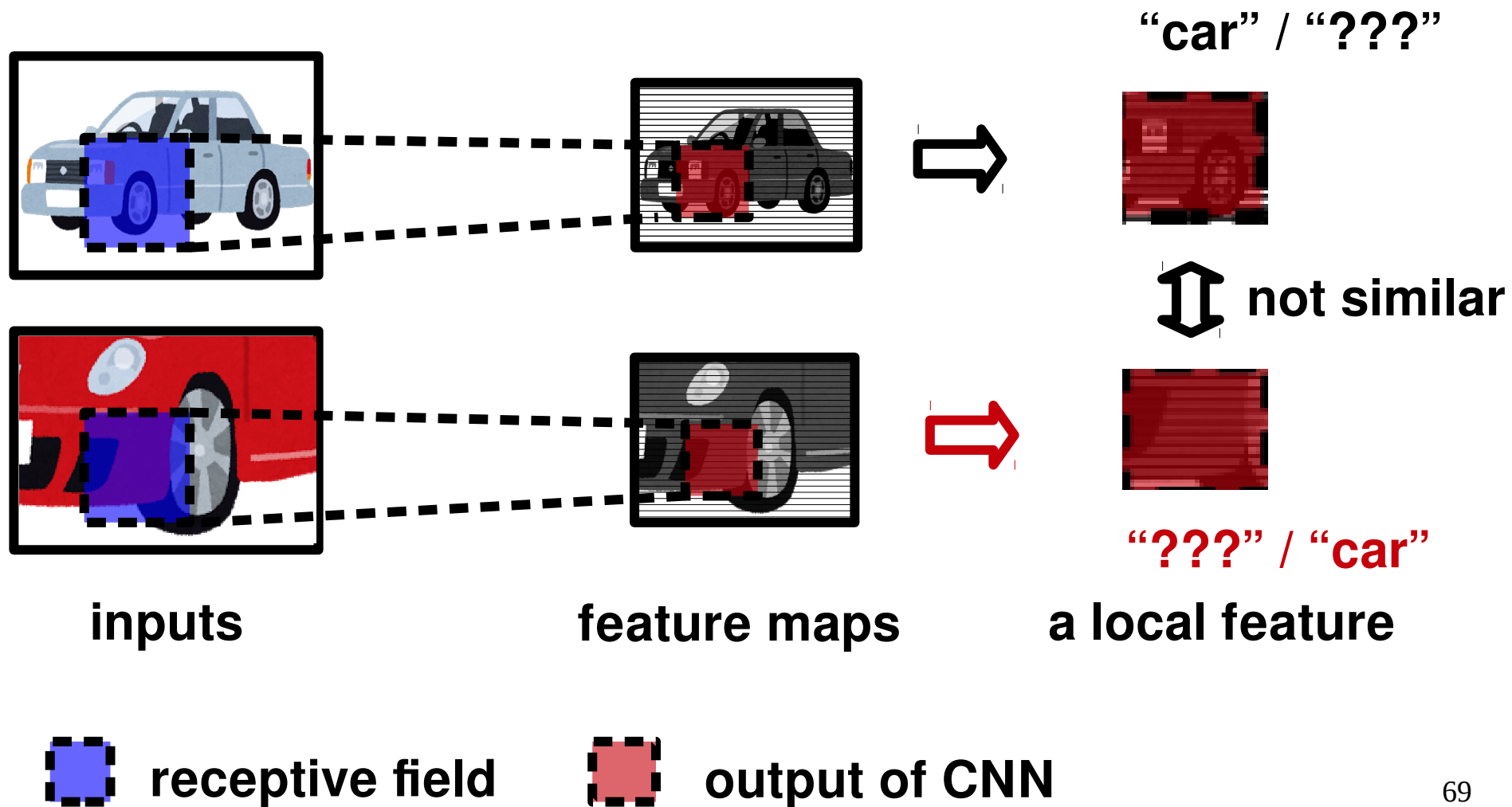
How the Scale Invariance?

► Multi-stage and sharing weights

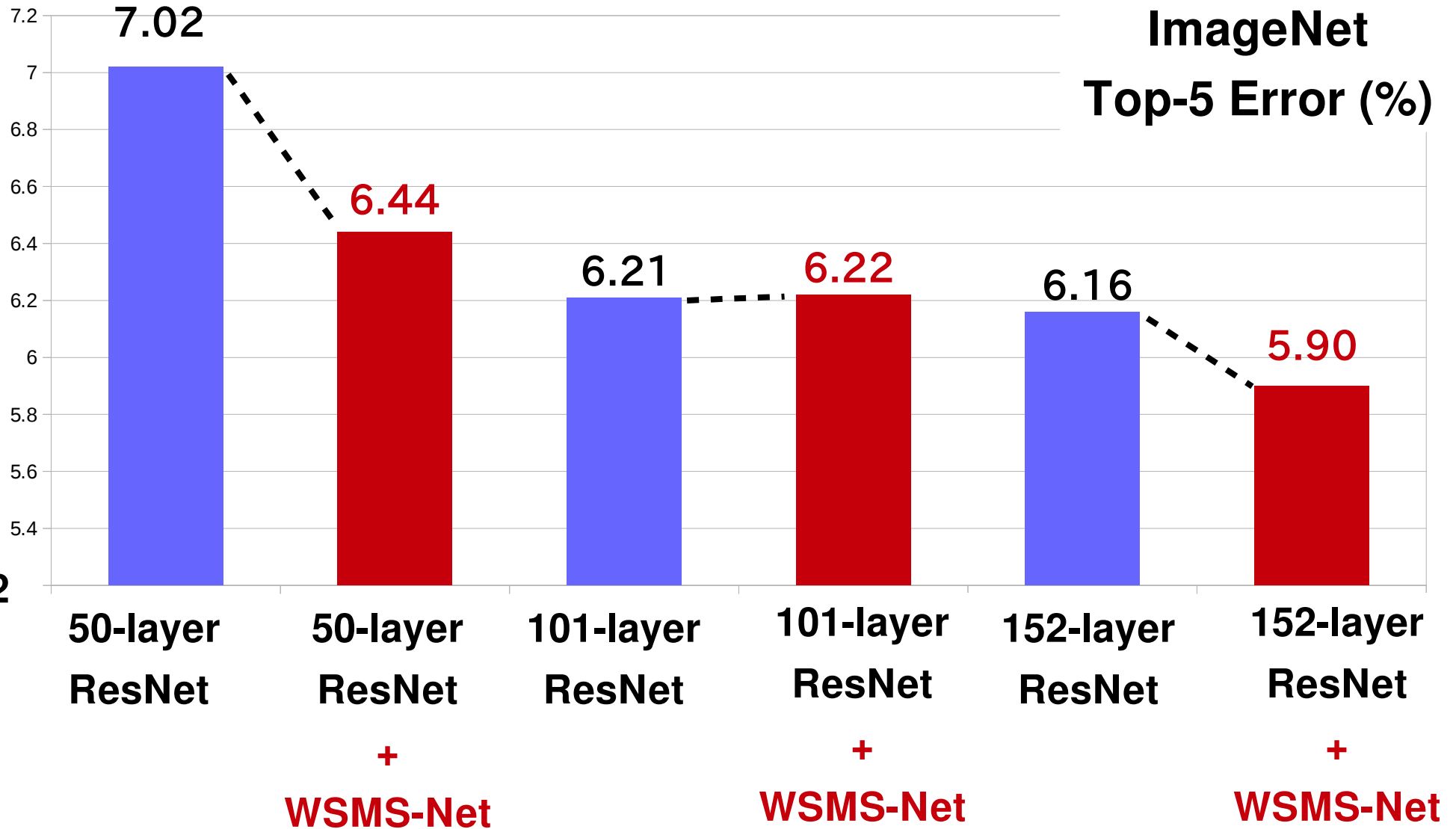


Our Approach

- ▶ **Receptive field** of CNN does not cover the same area



Experimental Results generalization capability

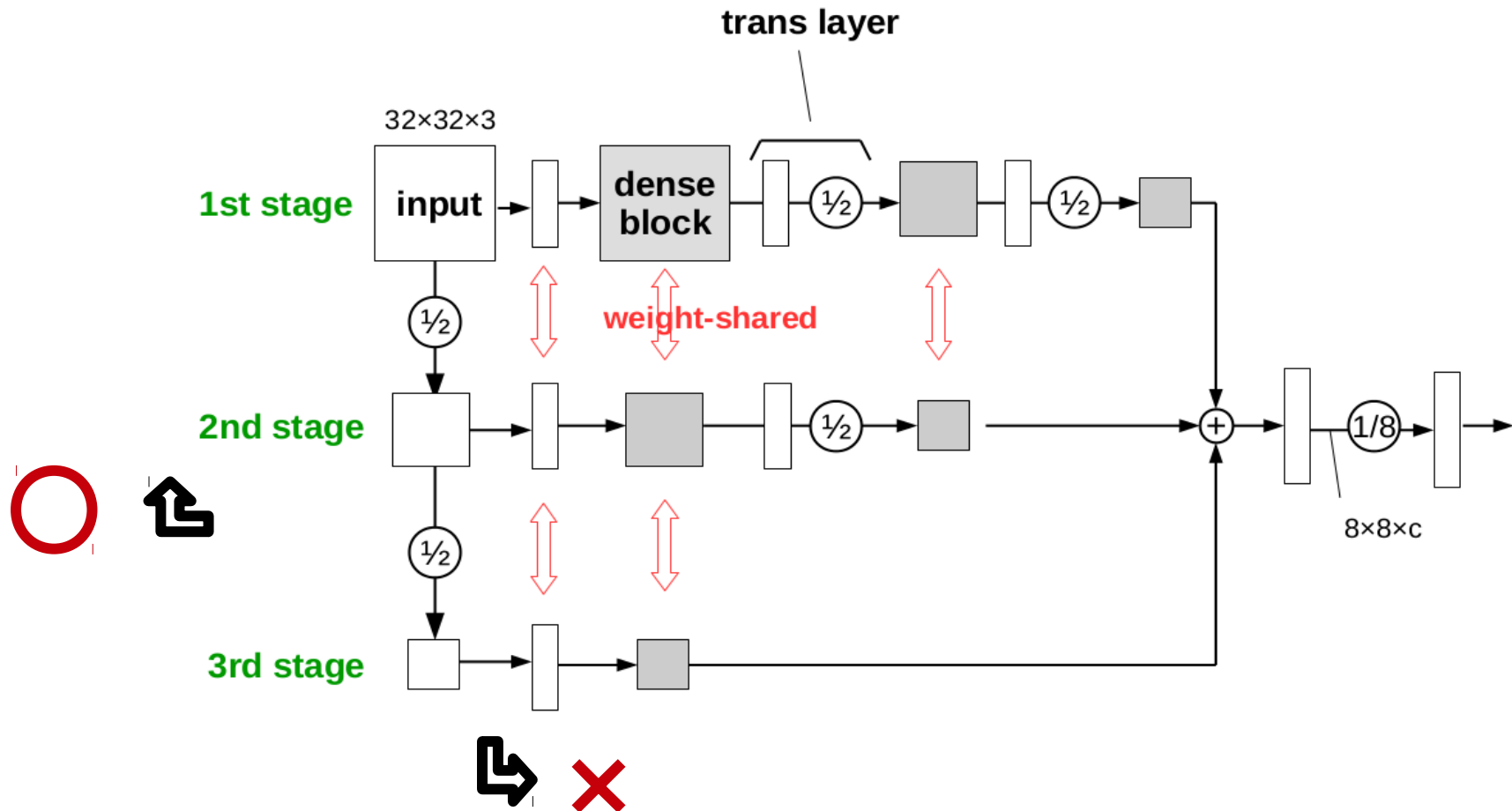


5.2

Experimental Results number of stages

► WSMS-Net with DenseNet for CIFAR

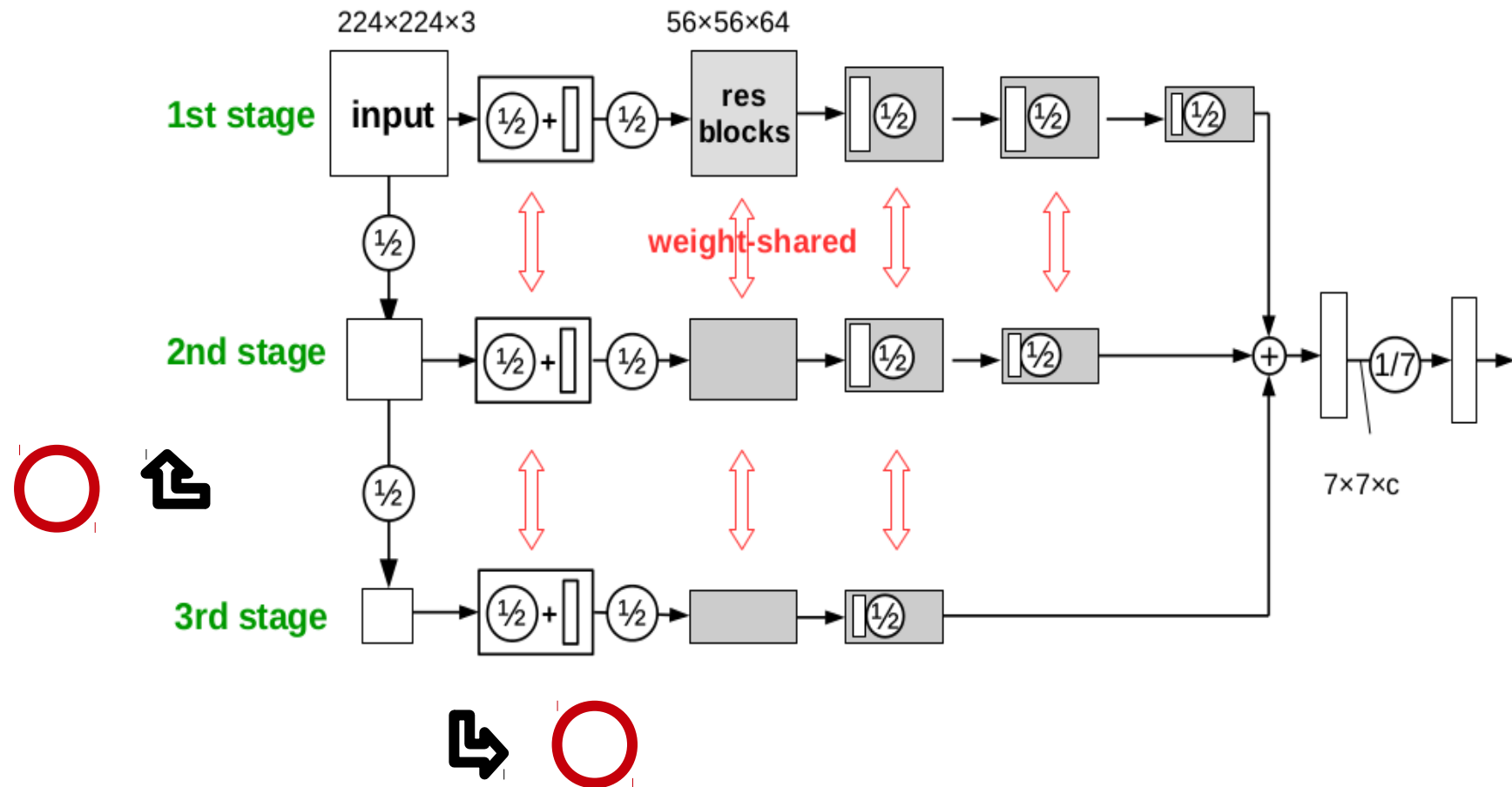
- 2 stages on the CIFAR
- 2 / 4 stages on the ImageNet



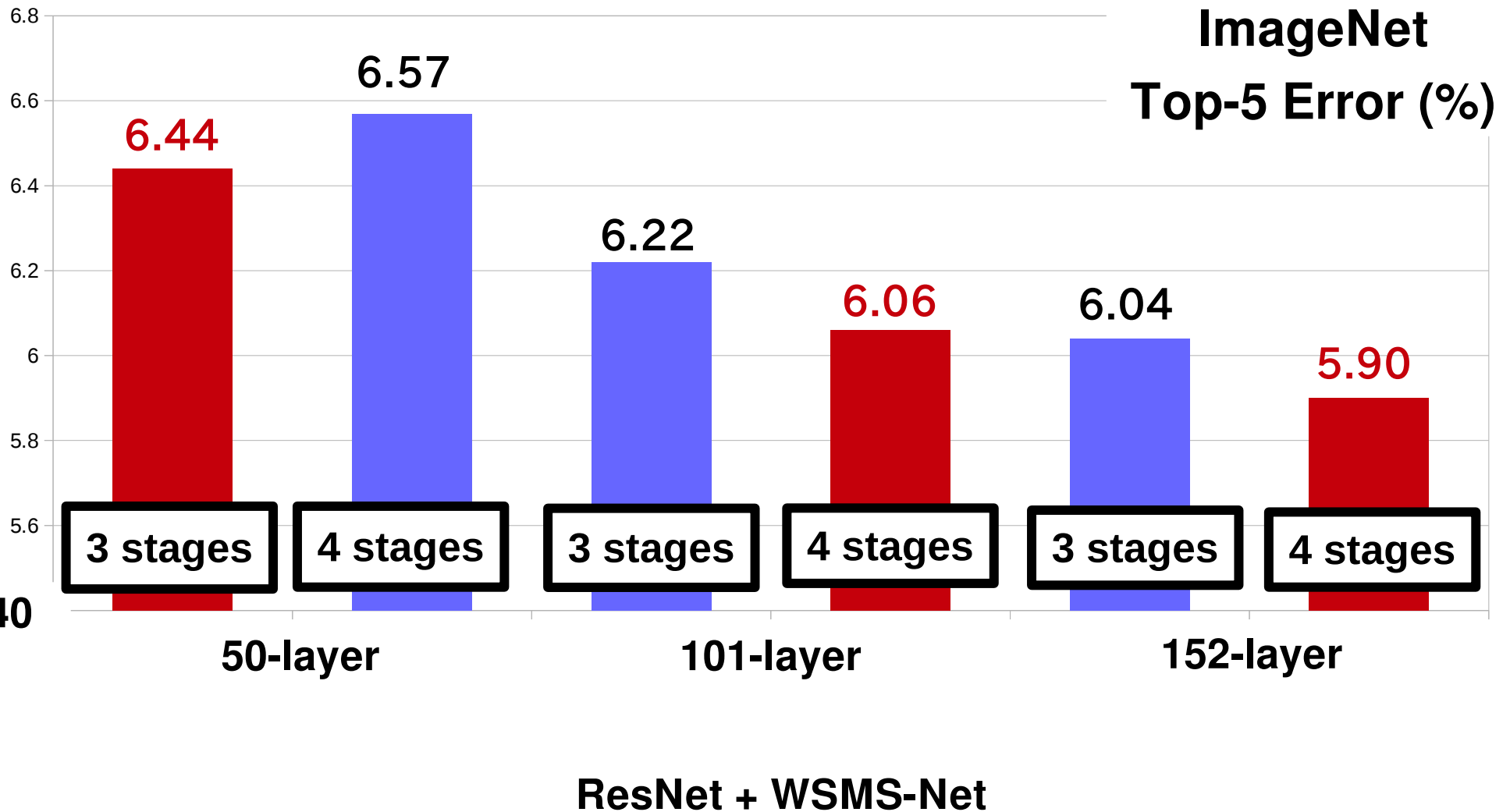
Experimental Results number of stages

► WSMS-Net with DenseNet for CIFAR

- 2 stages on the CIFAR
- **2 / 4 stages on the ImageNet**



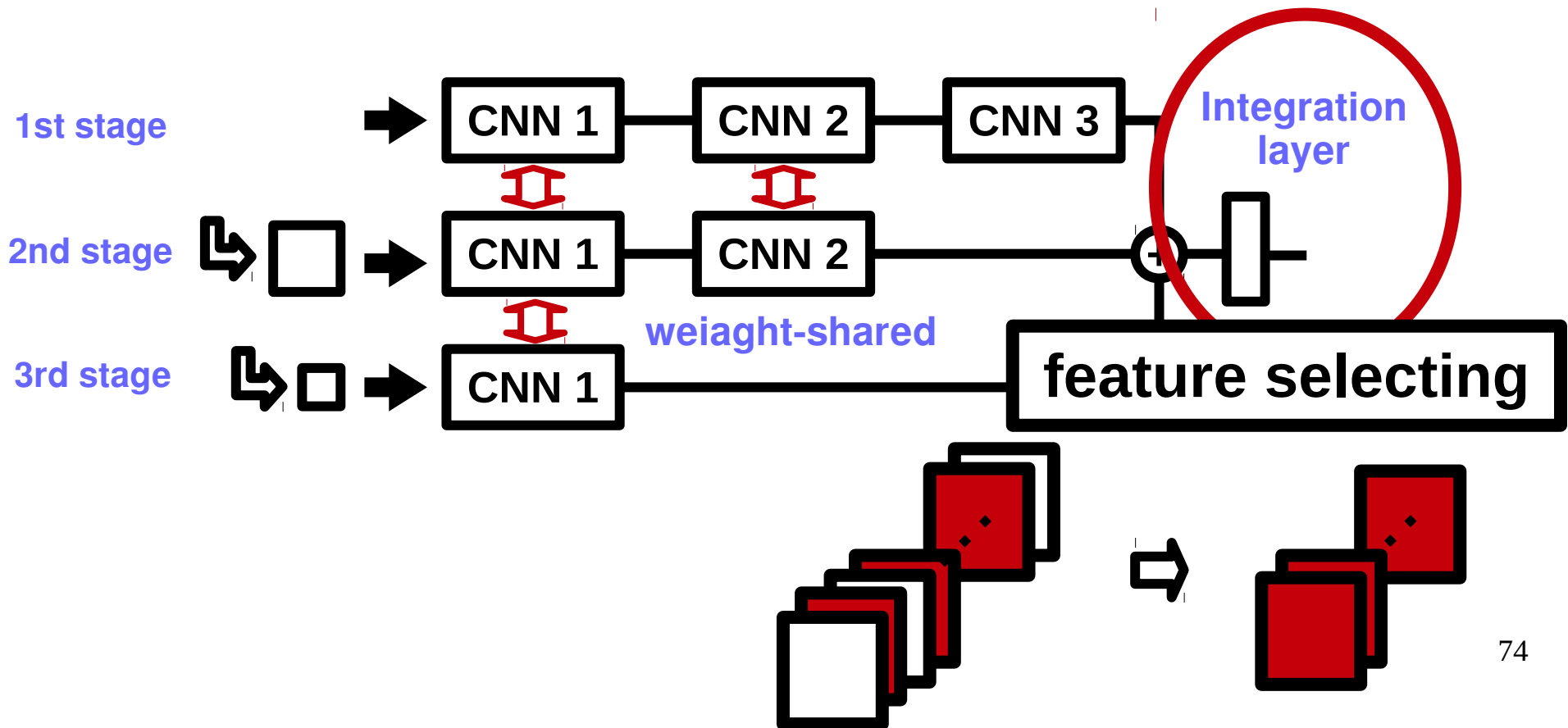
Experimental Results number of stages



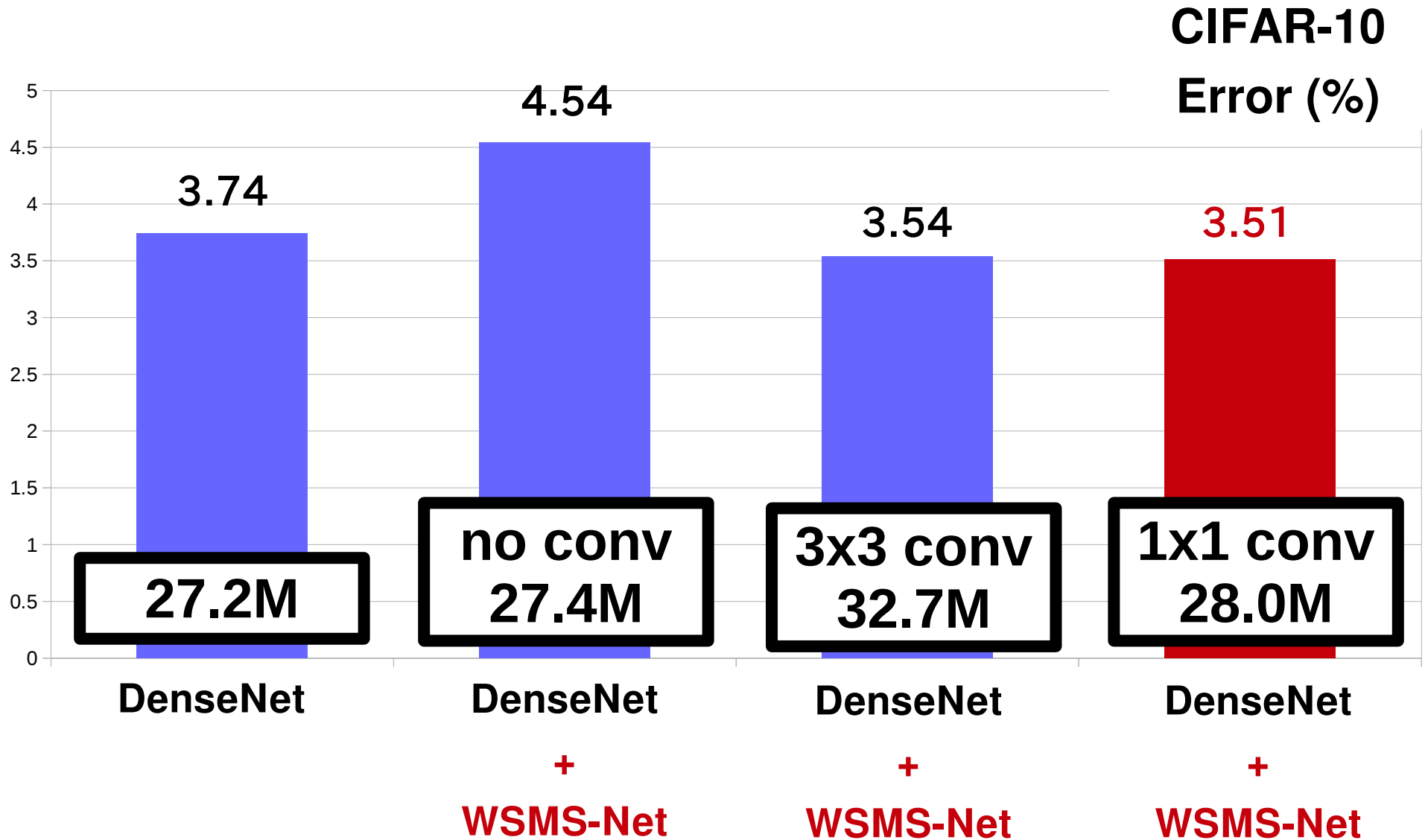
Experimental Results -detailed experiments-

► variations of **Integration Layer** (3 pattern)

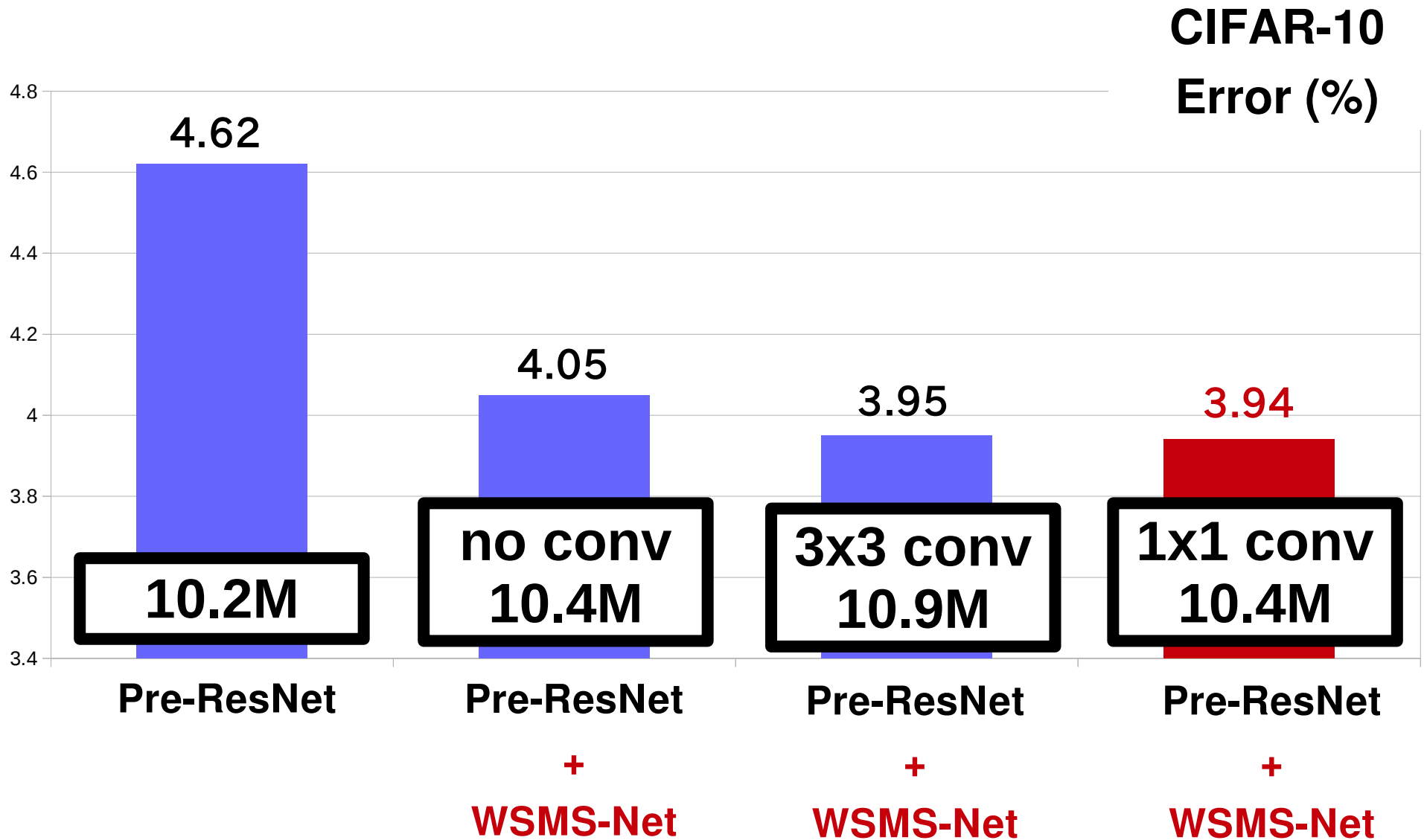
- use all feature maps from multiple stages (no conv)
- 3 x 3 convolution
- 1 x 1 convolution



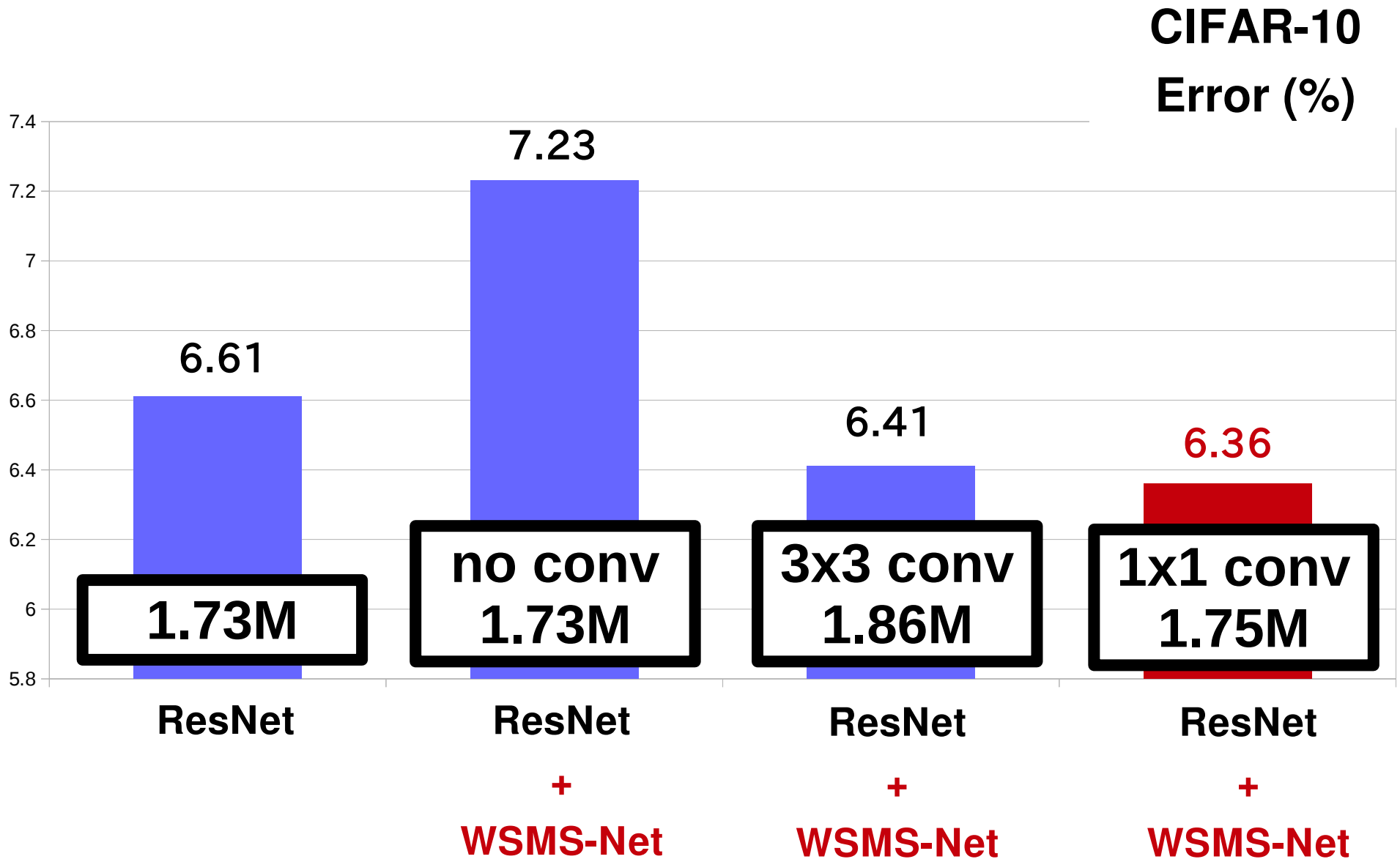
Experimental Results -detailed experiments-



Experimental Results -detailed experiments-

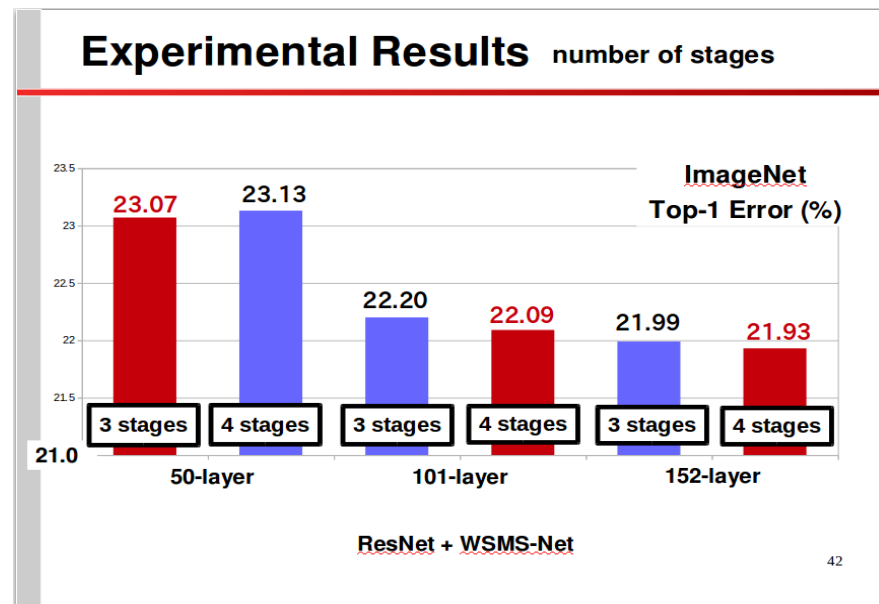
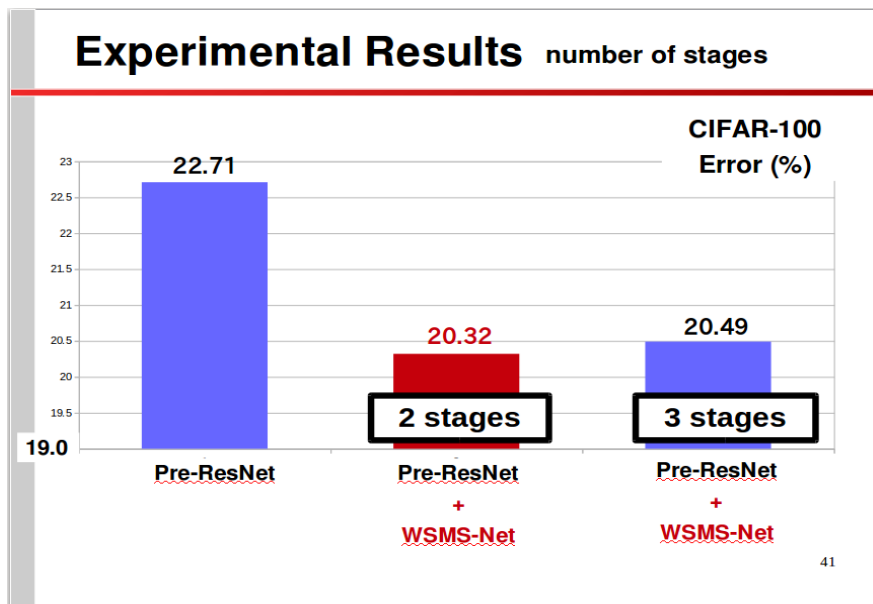
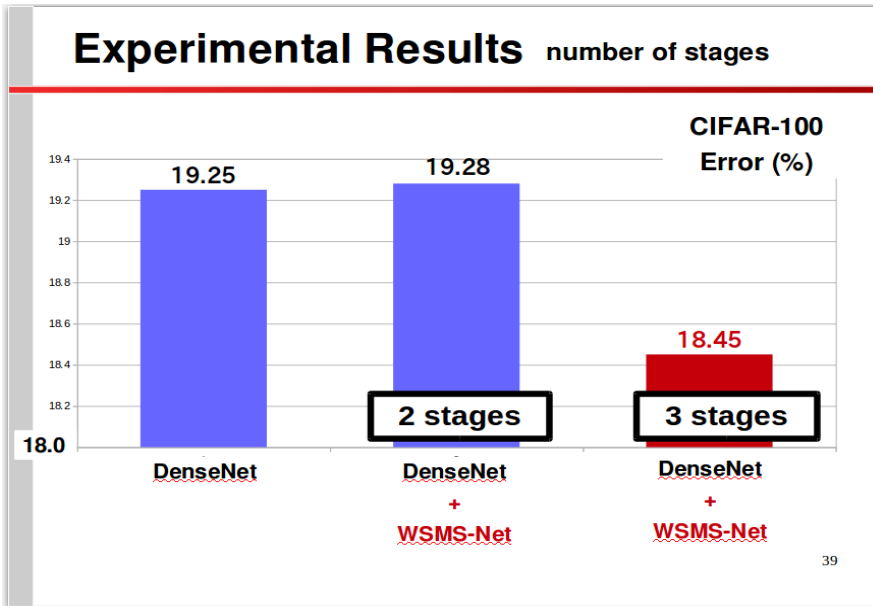


Experimental Results -detailed experiments-



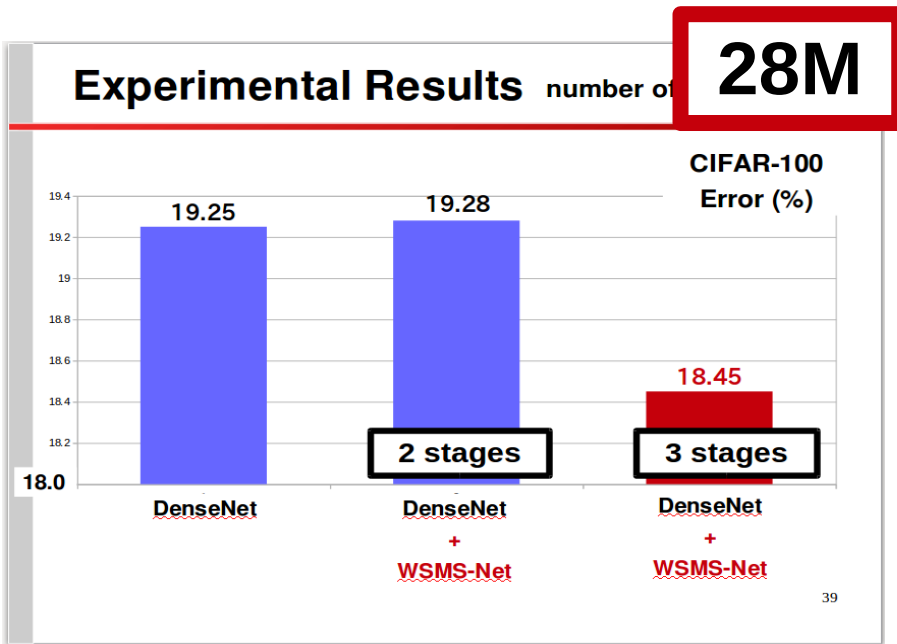
Future Works

number of stages

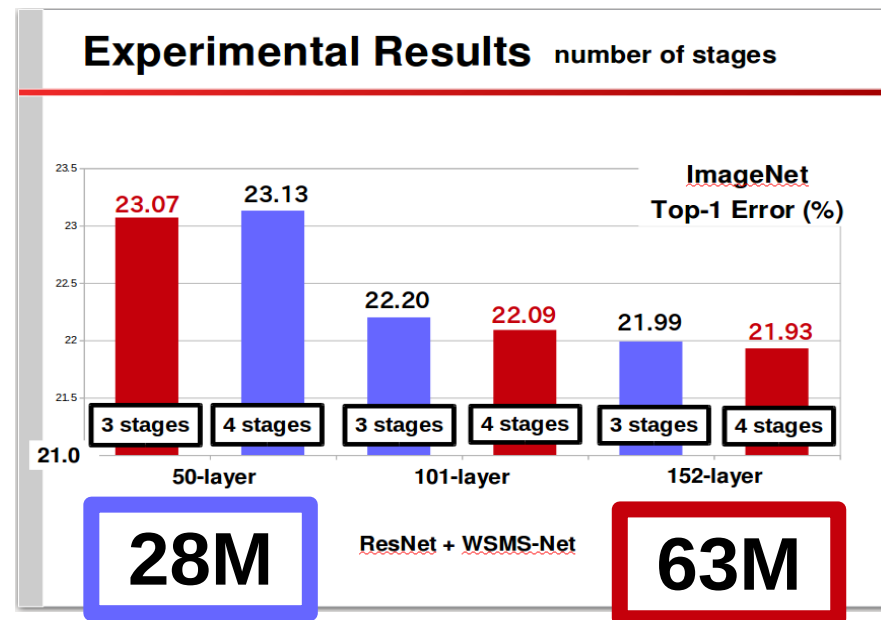
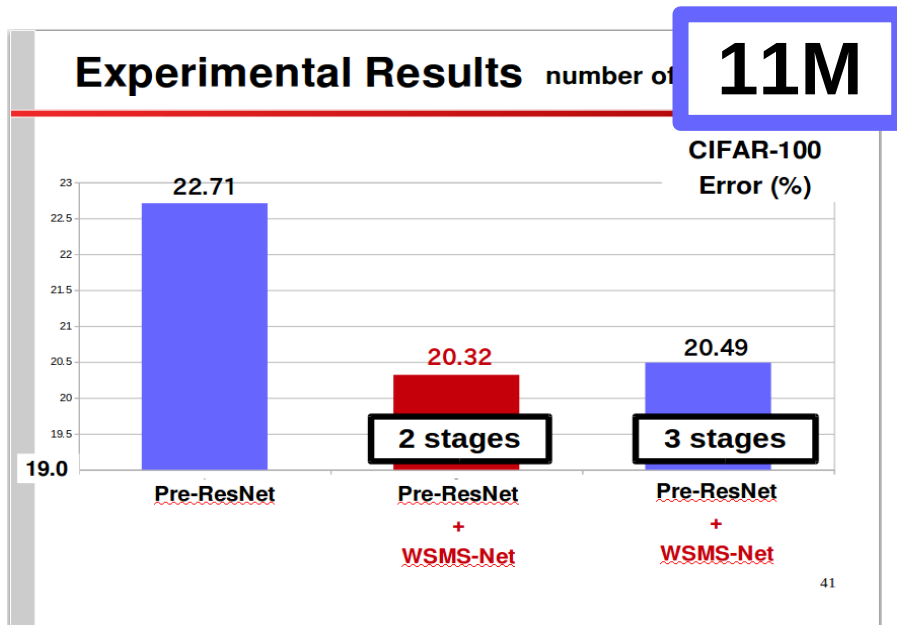
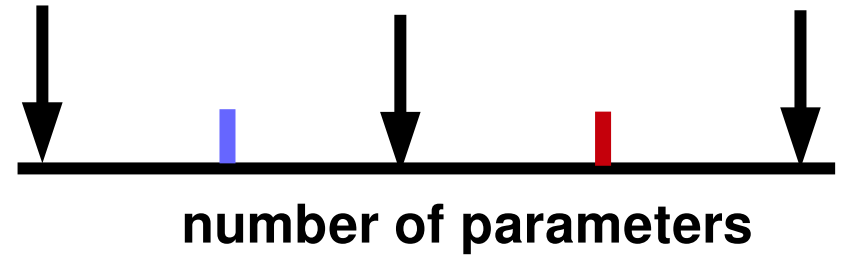


Future Works

number of stages

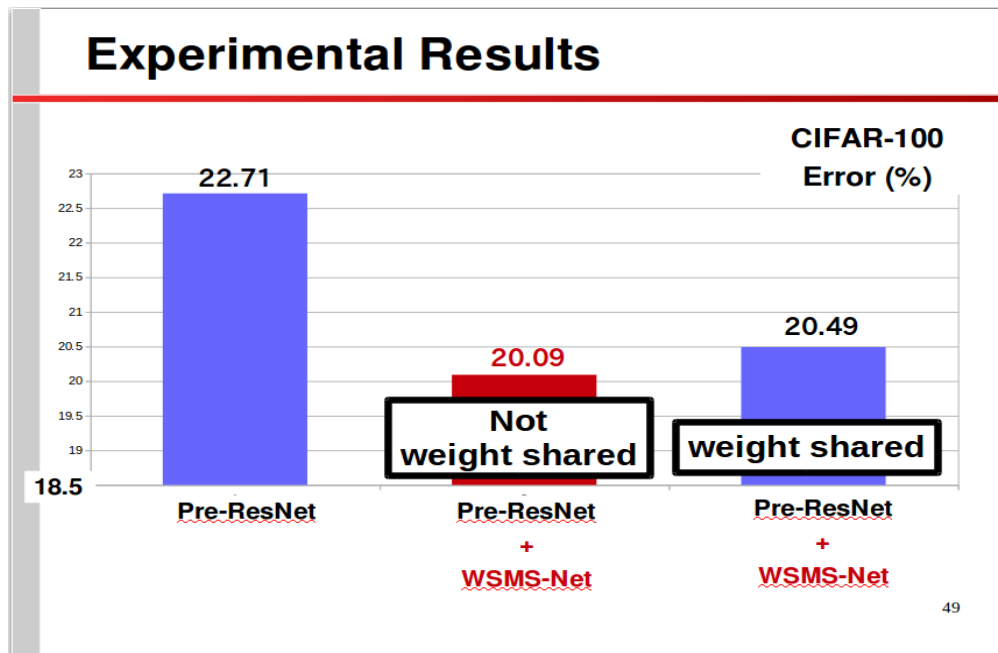


enough parameters : good effect
not enough parameters : bad effect



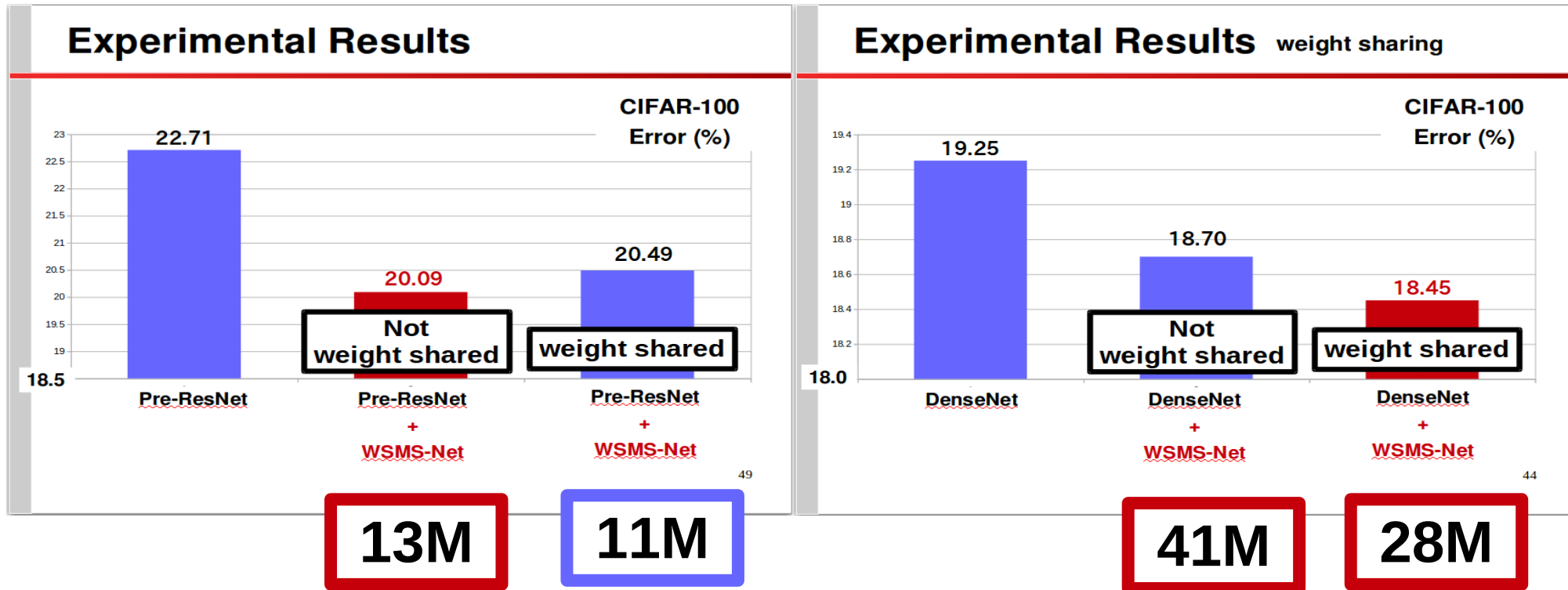
Future Works

weight sharing



Future Works

weight sharing



more parameters allow the model the more performance

