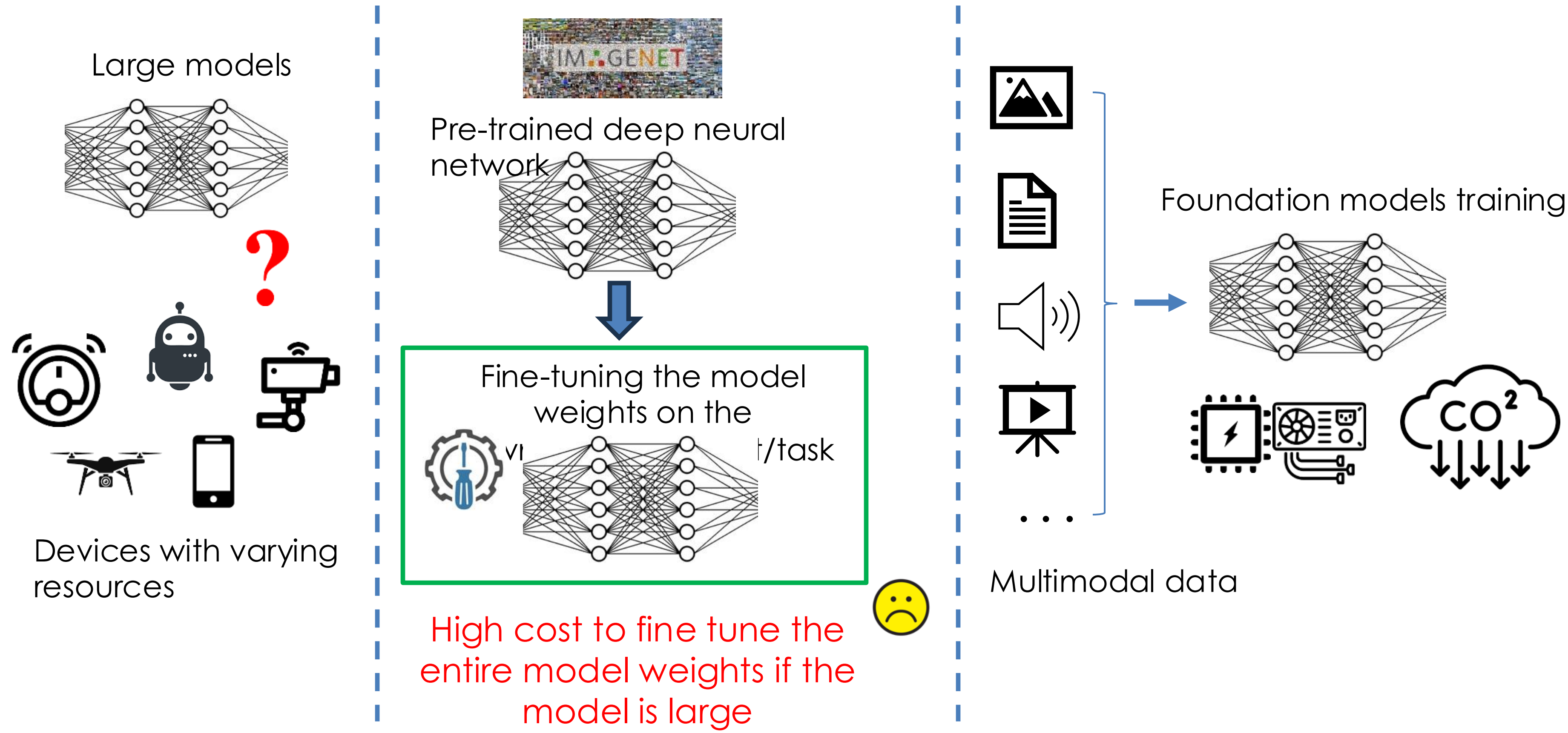
CAP 5516
Medical Image Computing
(Spring 2025)

Dr. Chen Chen
Associate Professor
Center for Research in Computer Vision (CRCV)
University of Central Florida
Office: HEC 221
Email: chen.chen@crcv.ucf.edu
Web: https://www.crcv.ucf.edu/chenchen/

UCF CENTER FOR RESEARCH IN COMPUTER VISION

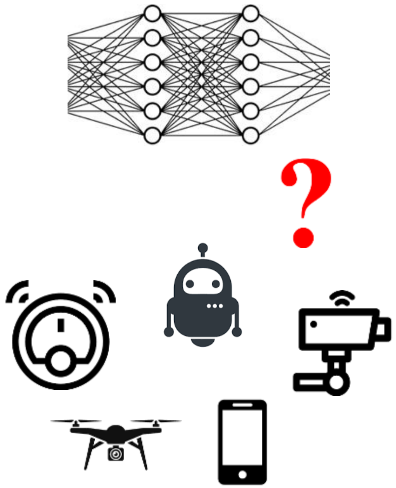
Lecture 12
Efficient Deep Learning (2)

Research Challenges



Research Challenges

Large models

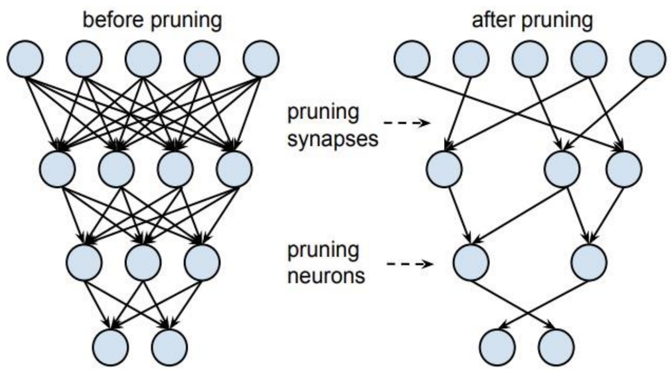


Devices with varying resources

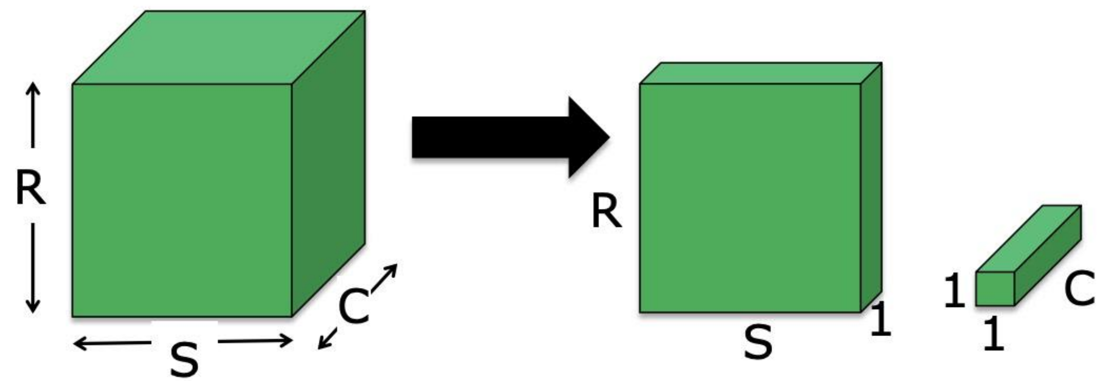
Efficient Neural Networks Design

Credit: Vivienne Sze

Network Pruning



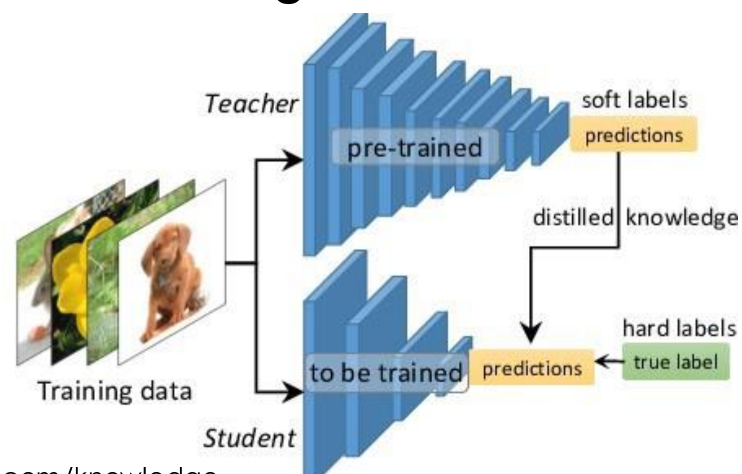
Efficient Network Architectures



[MobileNets, ShuffleNets, AdderNet]

Reduce Precision

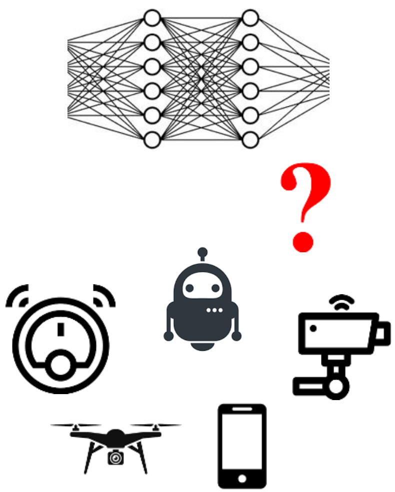
Knowledge Distillation



Source: https://towardsdatascience.com/knowledge-distillation-simplified-dd4973dbc764

Is Efficient Network the Ultimate Solution?

Large models



Devices with varying resources

Adaptive/Dynamic Network for Image Understanding

Taojiannan Yang, Sijie Zhu, Chen Chen, Shen Yan, Mi Zhang, Andrew Willis. "Mutualnet: Adaptive convnet via mutual learning from network width and resolution." European Conference on Computer Vision (ECCV), 2020. Oral Presentation.

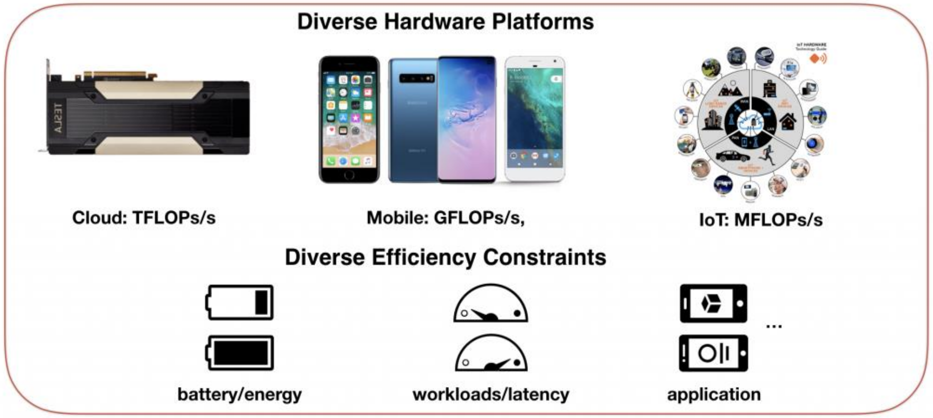
Yang, Taojiannan, Sijie Zhu, Matias Mendieta, Pu Wang, Ravikumar Balakrishnan, Minwoo Lee, Tao Han, Mubarak Shah, and Chen Chen. "MutualNet: Adaptive convnet via mutual learning from different model configurations." IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), Volume: 45, Issue: 1, 01 January 2023.

Research Problem

- Traditional Neural Networks are Static

	MobileNet	ResNet-50	ViT-B/16
Params	4.2M	25.6M	86M
FLOPs	575M	4.1G	17.5G

FLOPs: number of floating-point operations



Traditional neural networks (even efficient networks) are only executable at a specific resource constraint.

In real-world applications, **resource budgets are always changing** with many conditions (e.g., hardware, battery, task priority, etc.).

How to cope with dynamic resources and achieve a trade-off between accuracy and efficiency at inference?

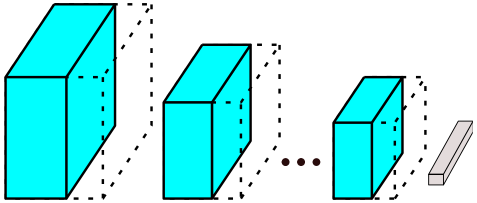
Research Problem

How to cope with dynamic resources and achieve a trade-off between accuracy and efficiency at inference?

Reducing complexity by width without retraining the network, the performance drops dramatically.

Width	1.0×		0.75×		0.5×	
re-train	✓	✗	✓	✗	✓	✗
Acc (%)	70.6	70.6	68.4	14.2	63.3	0.4

Neural Network: MobileNet



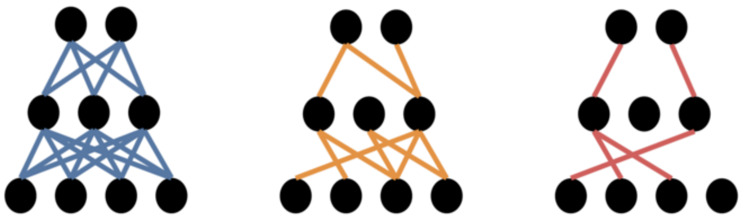
Reducing network width

Research Problem

How to cope with dynamic resources and achieve a trade-off between accuracy and efficiency at inference?

- One possible solution: install all the possible model variants with various resource-accuracy trade-offs in the heterogeneous AI systems
  - Consumes more memory and storage
  - Not scalable

Pruned networks with various pruning ratios



Different models with different sizes

Model	Params	FLOPs
ResNet-50	25.5M	4.1G
MobileNet v1	4.2M	569M
MobileNet v2	3.5M	300M

Motivation

The computational cost of a vanilla convolution = $C_{in} \times C_{out} \times K \times K \times H \times W$

K is the kernel size

C_{in} *and* C_{out} are the number of input and output channels

Related to the network size

H and W are output feature map sizes

Related to the input image size

Tuning knobs for computational cost

Motivation

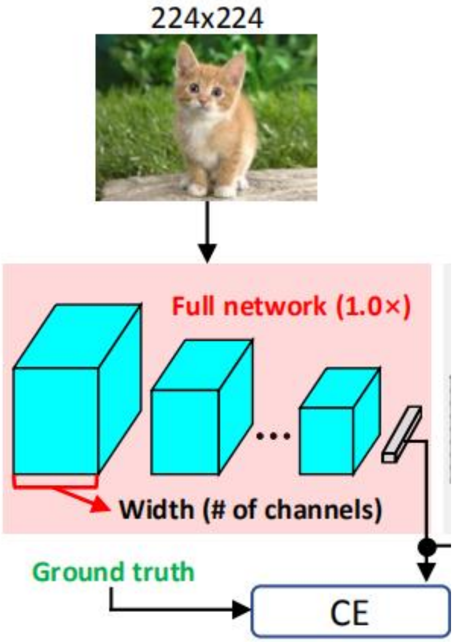
The computational cost of a vanilla convolution = $C_{in} \times C_{out} \times K \times K \times H \times W$

C_{in} *and* C_{out} are the number of input and output channels, K is the kernel size, H and W are output feature map sizes.

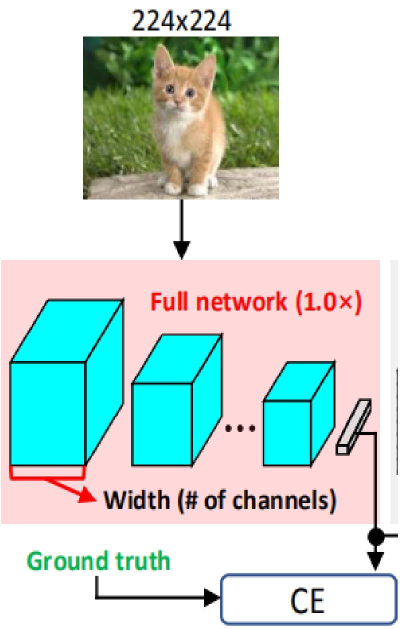
Objective:

Balancing between network width and input resolution to achieve a good accuracy-efficiency tradeoff at runtime with **a single network**.

Method (MutualNet Training)

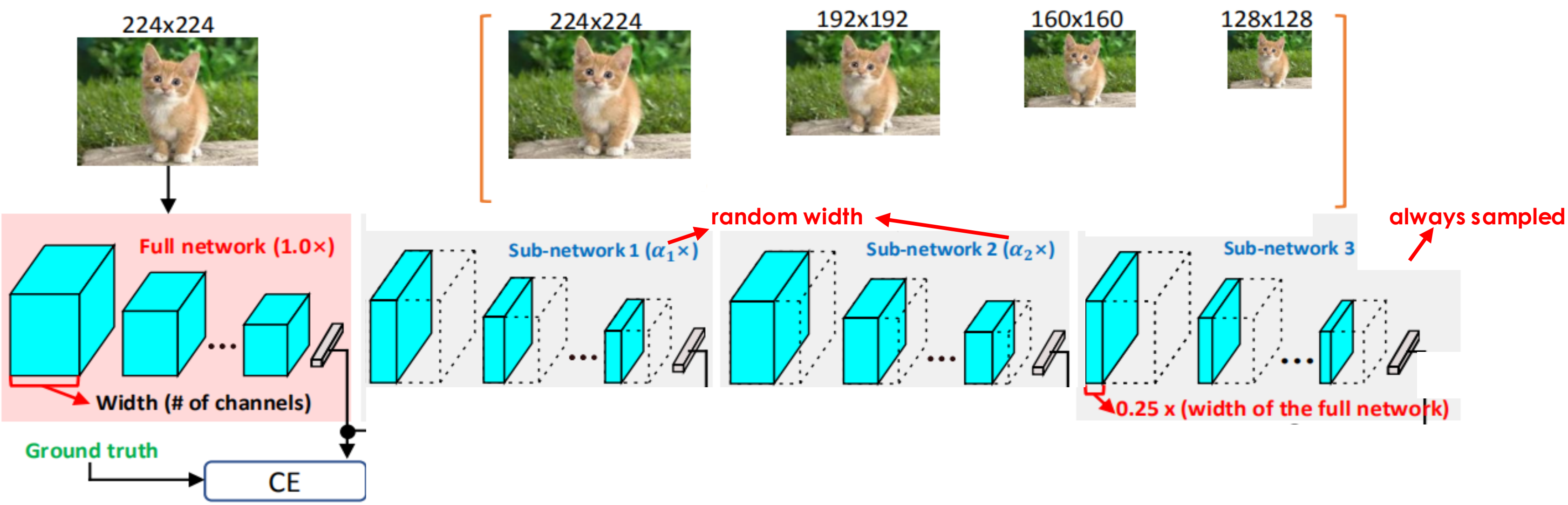


Method (MutualNet Training)

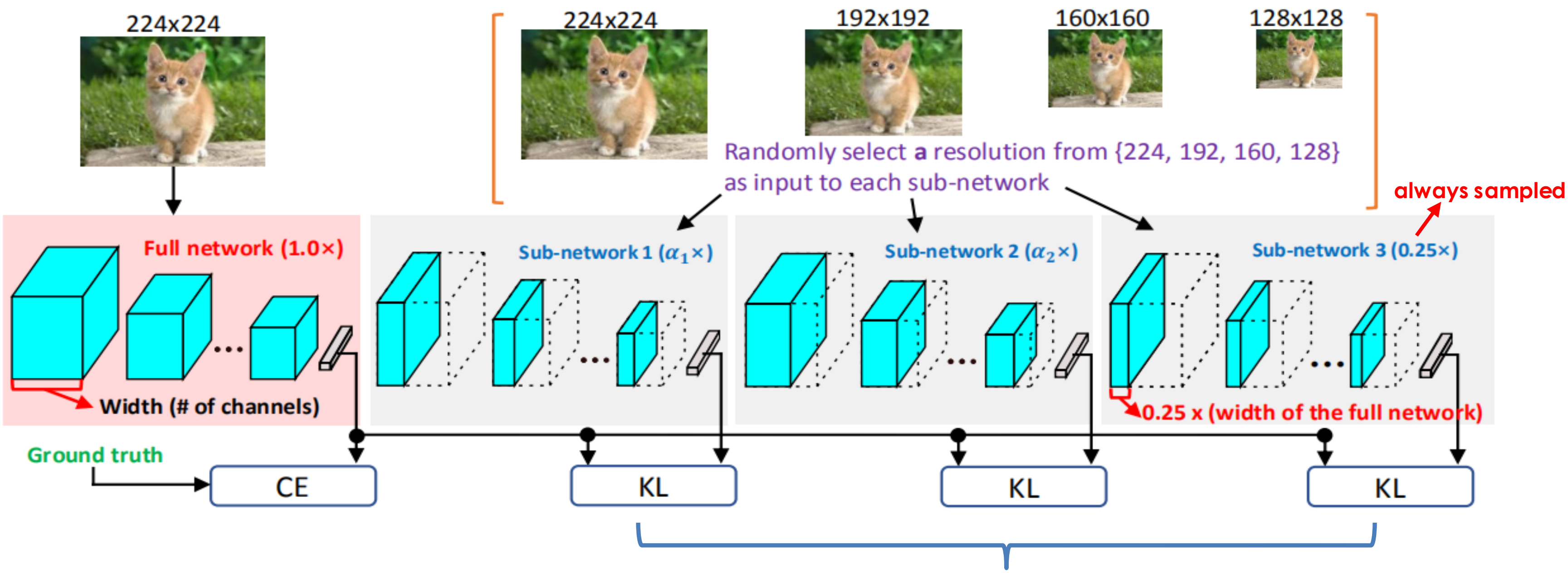


UCF CENTER FOR RESEARCH IN COMPUTER VISION

Method (MutualNet Training)



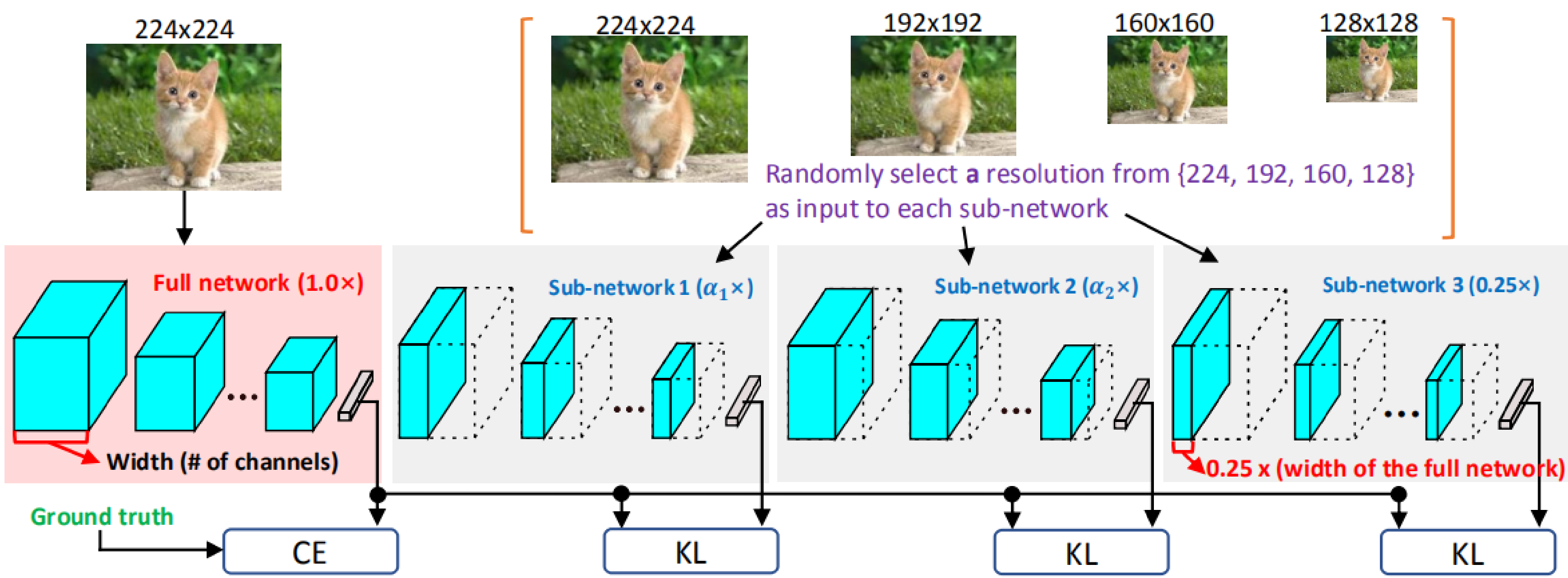
Method (MutualNet Training)



Weight sharing among all the networks

Knowledge distillation

Method (MutualNet Training)

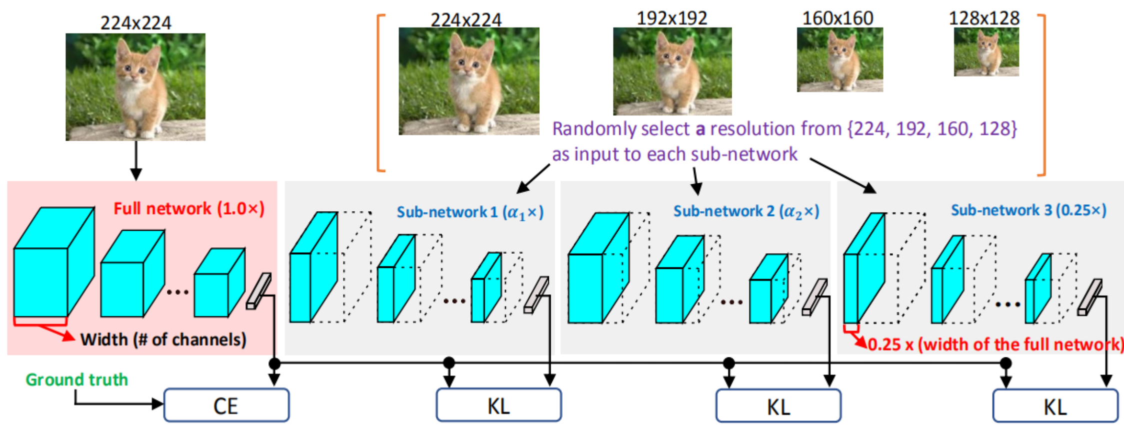


$$loss = loss_{full} + \sum_{i=1}^{3} loss_{sub_i}$$

UCF Center for Research in Computer Vision

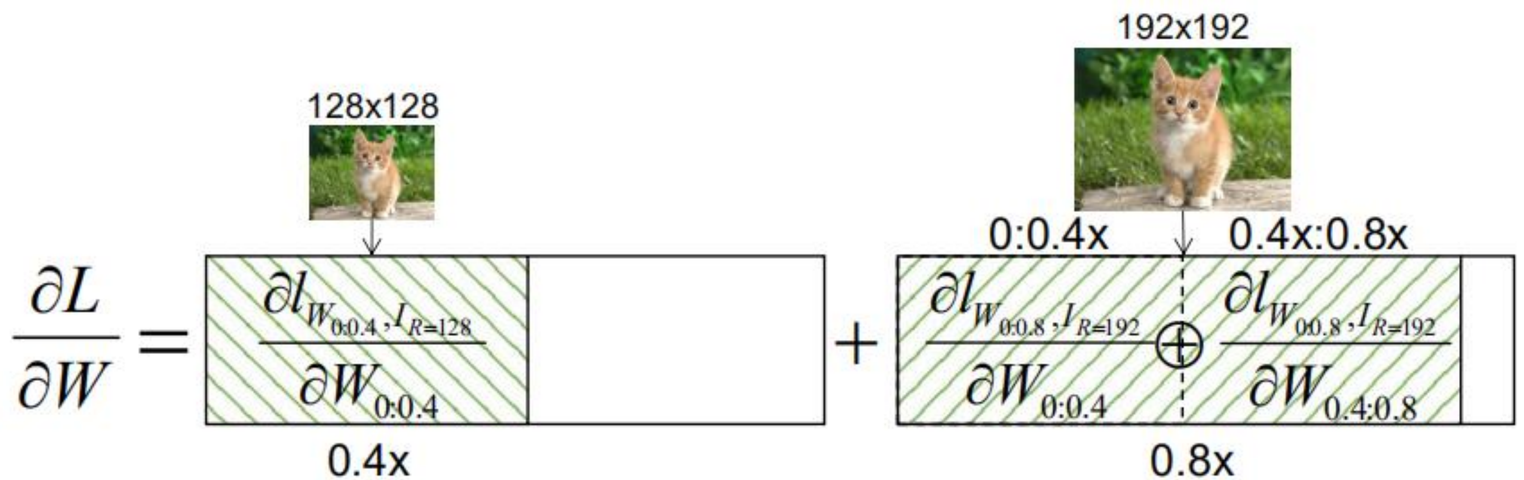
Method (MutualNet Training)

- The mutual learning scheme involves collaborative learning among an ensemble of networks.
- Sub-networks share weights and optimize together, enabling knowledge transfer.



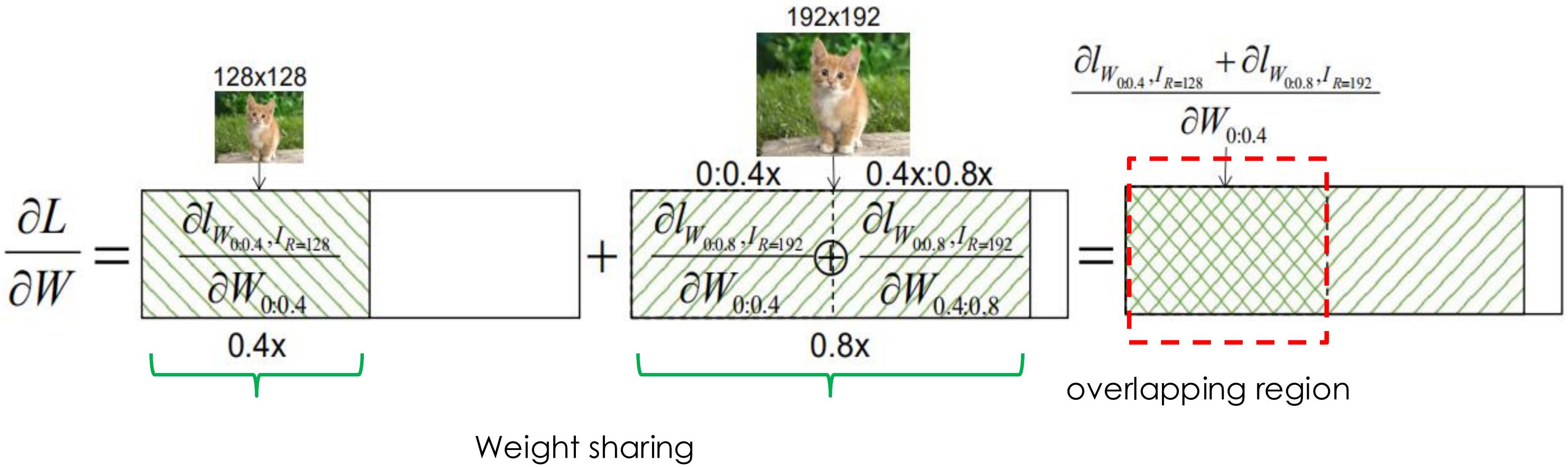
Gradient Analysis

- Consider two network widths 0.4× and 0.8×, and two resolutions 128 and 192 as an example



Gradient Analysis

- Consider two network widths 0.4× and 0.8×, and two resolutions 128 and 192 as an example

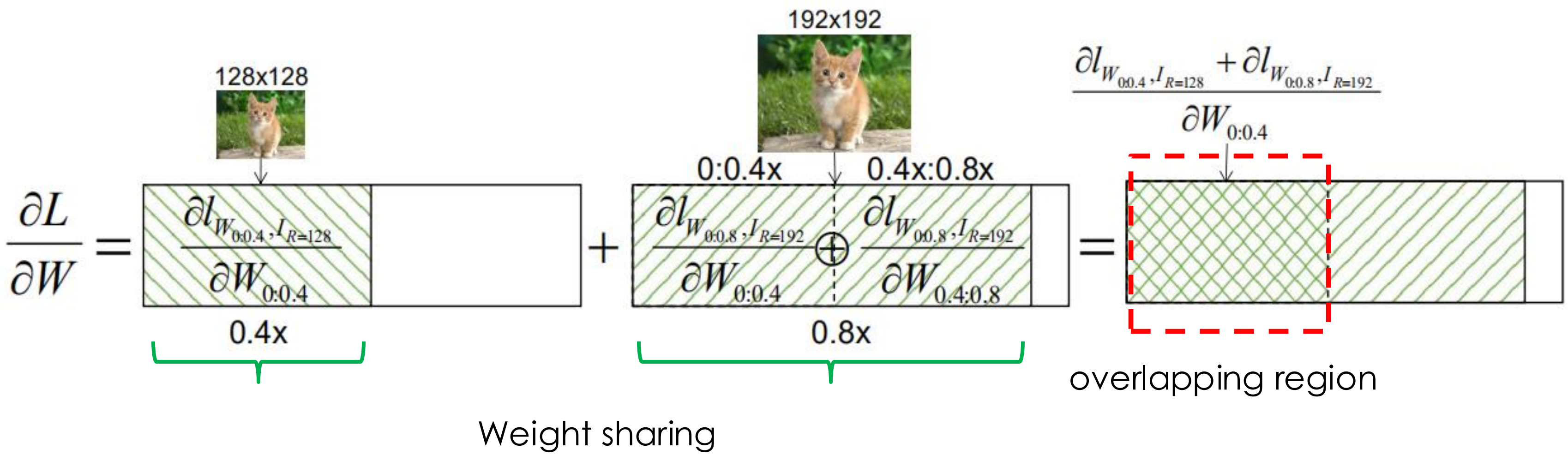


The gradients of sub-network 0.4× is

$$\frac{\partial l_{W_{0:0.4}, I_{R=128}} + \partial l_{W_{0:0.8}, I_{R=192}}}{\partial W_{0:0.4}}$$

Gradient Analysis

- Consider two network widths 0.4× and 0.8×, and two resolutions 128 and 192 as an example



The gradients of sub-network 0.4× is

$$\frac{\partial l_{W_{0:0.4},I_{R=128}} + \partial l_{W_{0:0.8},I_{R=192}}}{\partial W_{0:0.4}}$$

Multi-scale = Network scale + Input scale

Method (MutualNet Inference)

After training, we test the performance of different width-resolution configurations on a Validation Set.

Top1 Acc

reso/width	1.0x	0.95x	0.9x	0.85x	0.8x	0.75x	0.7x	0.65x	0.6x	0.55x	0.5x	0.45x	0.4x	0.35x	0.3x	0.25x
224	72.4	71.7	71.1	70.4	69.8	69.1	68.3	67.2	66	64.6	63.1	61.6	59.5	57.2	55.3	53.6
192	70.9	70.6	70.2	69.7	69.1	68.4	67.5	66.7	65.5	63.8	62.2	60.8	58.5	56.6	54.6	52.7
160	68.6	68.5	68.1	67.7	67.2	66.5	65.6	64.7	63.5	61.8	60.3	58.9	56.6	54.4	52.4	50.1
128	64	64	64	63.8	63.1	62.5	61.6	60.6	59.5	57.6	56.1	54.3	52.1	50.1	47.7	45.5

MFLOPs

reso/width	1.0x	0.95x	0.9x	0.85x	0.8x	0.75x	0.7x	0.65x	0.6x	0.55x	0.5x	0.45x	0.4x	0.35x	0.3x	0.25x
224	569	518	466	421	366	325	287	249	217	177	149	124	100	80	64	41
192	418	380	342	309	269	239	211	183	159	130	109	91	73	59	47	30
160	290	265	239	215	187	166	146	127	111	90	76	63	51	41	32	21
128	186	170	152	138	120	106	94	81	71	58	49	40	32	26	21	13

MobileNetv1 backbone

Method (MutualNet Inference)

After training, we test the performance of different width-resolution configurations on a Validation Set.

Choose the best one under a given constraint (FLOPs or latency).

Top1 Acc

reso/width	1.0x	0.95x	0.9x	0.85x	0.8x	0.75x	0.7x	0.65x	0.6x	0.55x	0.5x	0.45x	0.4x	0.35x	0.3x	0.25x
224	72.4	71.7	71.1	70.4	69.8	69.1	68.3	67.2	66	64.6	63.1	61.6	59.5	57.2	55.3	53.6
192	70.9	70.6	70.2	69.7	69.1	68.4	67.5	66.7	65.5	63.8	62.2	60.8	58.5	56.6	54.6	52.7
160	68.6	68.5	68.1	67.7	67.2	66.5	65.6	64.7	63.5	61.8	60.3	58.9	56.6	54.4	52.4	50.1
128	64	64	64	63.8	63.1	62.5	61.6	60.6	59.5	57.6	56.1	54.3	52.1	50.1	47.7	45.5

MFLOPs

reso/width	1.0x	0.95x	0.9x	0.85x	0.8x	0.75x	0.7x	0.65x	0.6x	0.55x	0.5x	0.45x	0.4x	0.35x	0.3x	0.25x
224	569	518	466	421	366	325	287	249	217	177	149	124	100	80	64	41
192	418	380	342	309	269	239	211	183	159	130	109	91	73	59	47	30
160	290	265	239	215	187	166	146	127	111	90	76	63	51	41	32	21
128	186	170	152	138	120	106	94	81	71	58	49	40	32	26	21	13

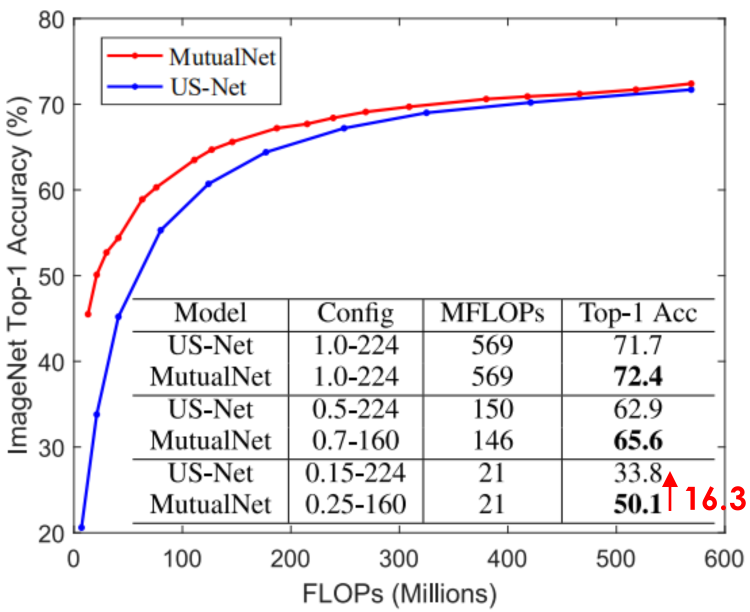
MobileNetv1 backbone

Method (MutualNet Inference)

For deployment, we only need to **deploy one model and the FLOPs-Acc query table**. Then we can adjust the model configuration according to different resource constraints.

Results – ImageNet Classification

ImageNet dataset: 1.2 million training images and 50,000 validation images in 1000 categories

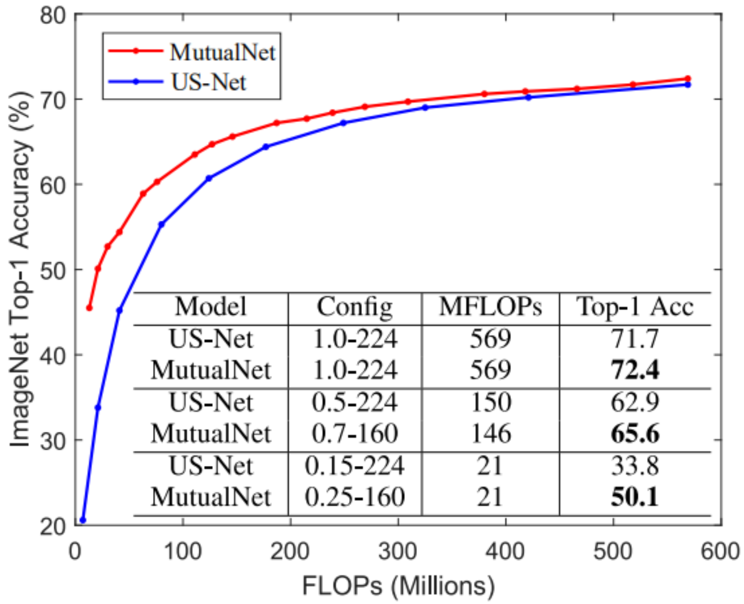


Model	Config	MFLOPs	Top-1 Acc
US-Net	1.0-224	569	71.7
MutualNet	1.0-224	569	**72.4**
US-Net	0.5-224	150	62.9
MutualNet	0.7-160	146	**65.6**
US-Net	0.15-224	21	33.8
MutualNet	0.25-160	21	**50.1**

MobileNet v1

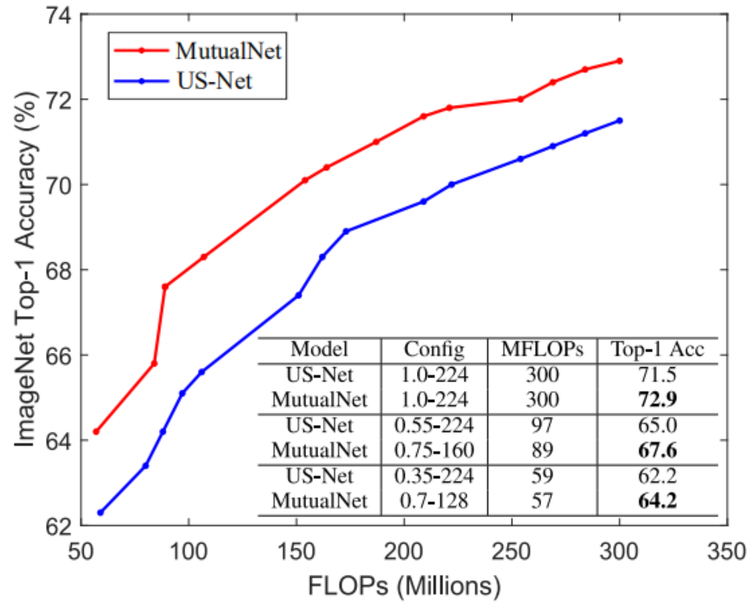
Results – ImageNet Classification

ImageNet dataset: 1.2 million training images and 50,000 validation images in 1000 categories



Model	Config	MFLOPs	Top-1 Acc
US-Net	1.0-224	569	71.7
MutualNet	1.0-224	569	**72.4**
US-Net	0.5-224	150	62.9
MutualNet	0.7-160	146	**65.6**
US-Net	0.15-224	21	33.8
MutualNet	0.25-160	21	**50.1**

MobileNet v1



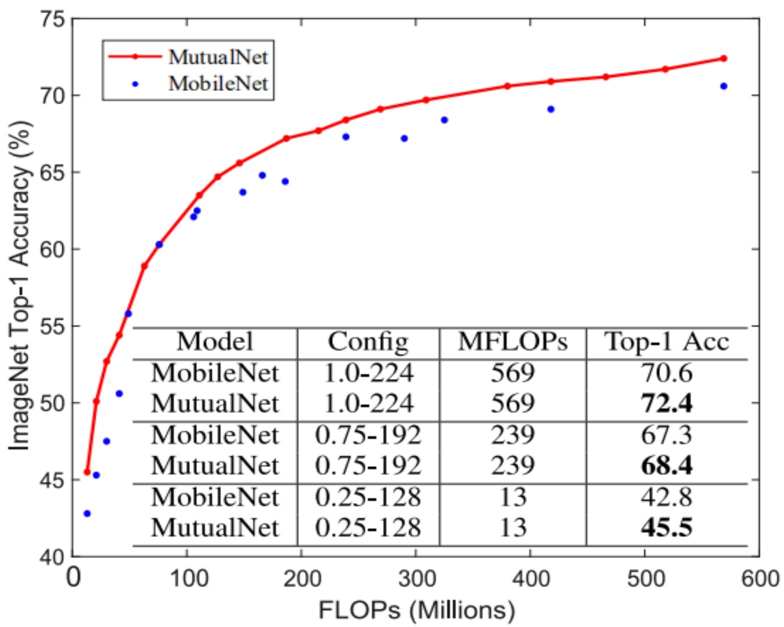
Model	Config	MFLOPs	Top-1 Acc
US-Net	1.0-224	300	71.5
MutualNet	1.0-224	300	**72.9**
US-Net	0.55-224	97	65.0
MutualNet	0.75-160	89	**67.6**
US-Net	0.35-224	59	62.2
MutualNet	0.7-128	57	**64.2**

MobileNet v2

Significantly outperforms state-of-the-art methods over the whole Acc-FLOPs spectrum

Results – ImageNet Classification

Comparison with individually trained networks

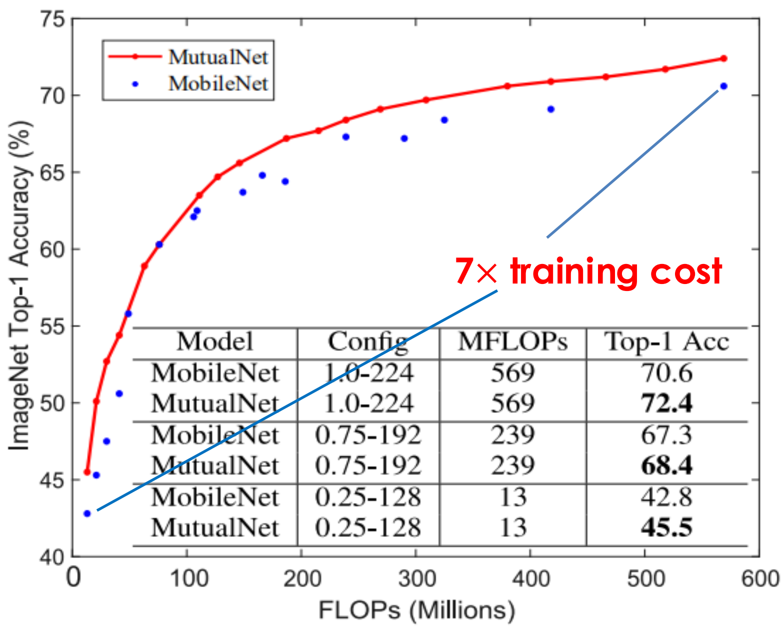


Model	Config	MFLOPs	Top-1 Acc
MobileNet	1.0-224	569	70.6
MutualNet	1.0-224	569	**72.4**
MobileNet	0.75-192	239	67.3
MutualNet	0.75-192	239	**68.4**
MobileNet	0.25-128	13	42.8
MutualNet	0.25-128	13	**45.5**

MobileNet v1

Results – ImageNet Classification

Comparison with individually trained networks



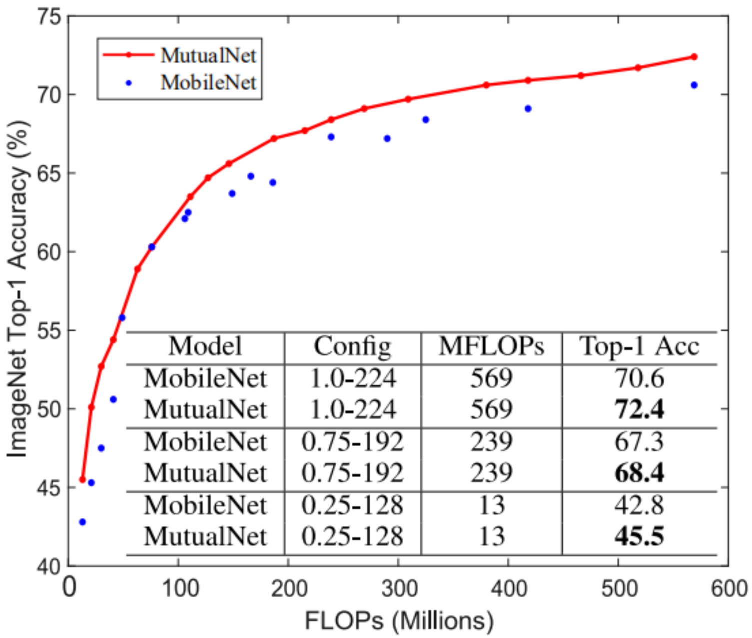
Model	Config	MFLOPs	Top-1 Acc
MobileNet	1.0-224	569	70.6
MutualNet	1.0-224	569	**72.4**
MobileNet	0.75-192	239	67.3
MutualNet	0.75-192	239	**68.4**
MobileNet	0.25-128	13	42.8
MutualNet	0.25-128	13	**45.5**

MobileNet v1

Training multiple networks will significantly increase the training cost

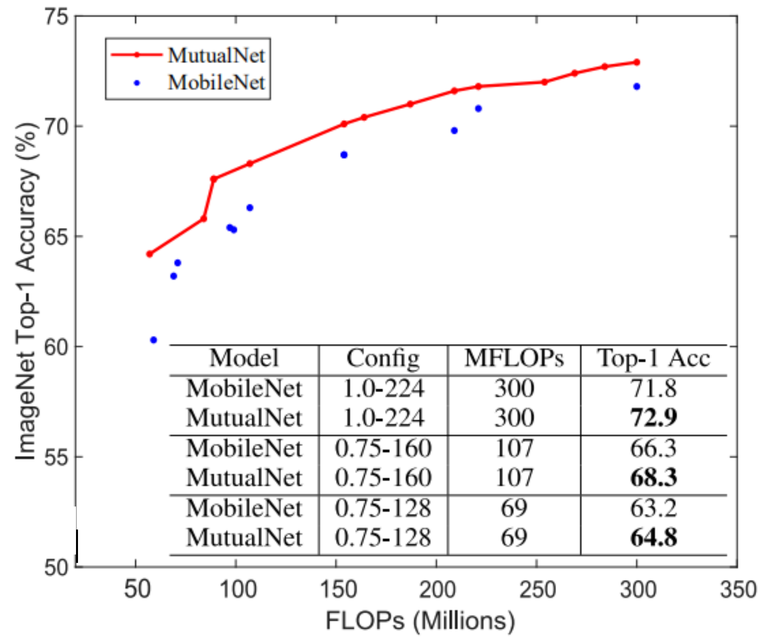
Results – ImageNet Classification

Comparison with individually trained networks



Model	Config	MFLOPs	Top-1 Acc
MobileNet	1.0-224	569	70.6
MutualNet	1.0-224	569	**72.4**
MobileNet	0.75-192	239	67.3
MutualNet	0.75-192	239	**68.4**
MobileNet	0.25-128	13	42.8
MutualNet	0.25-128	13	**45.5**

MobileNet v1

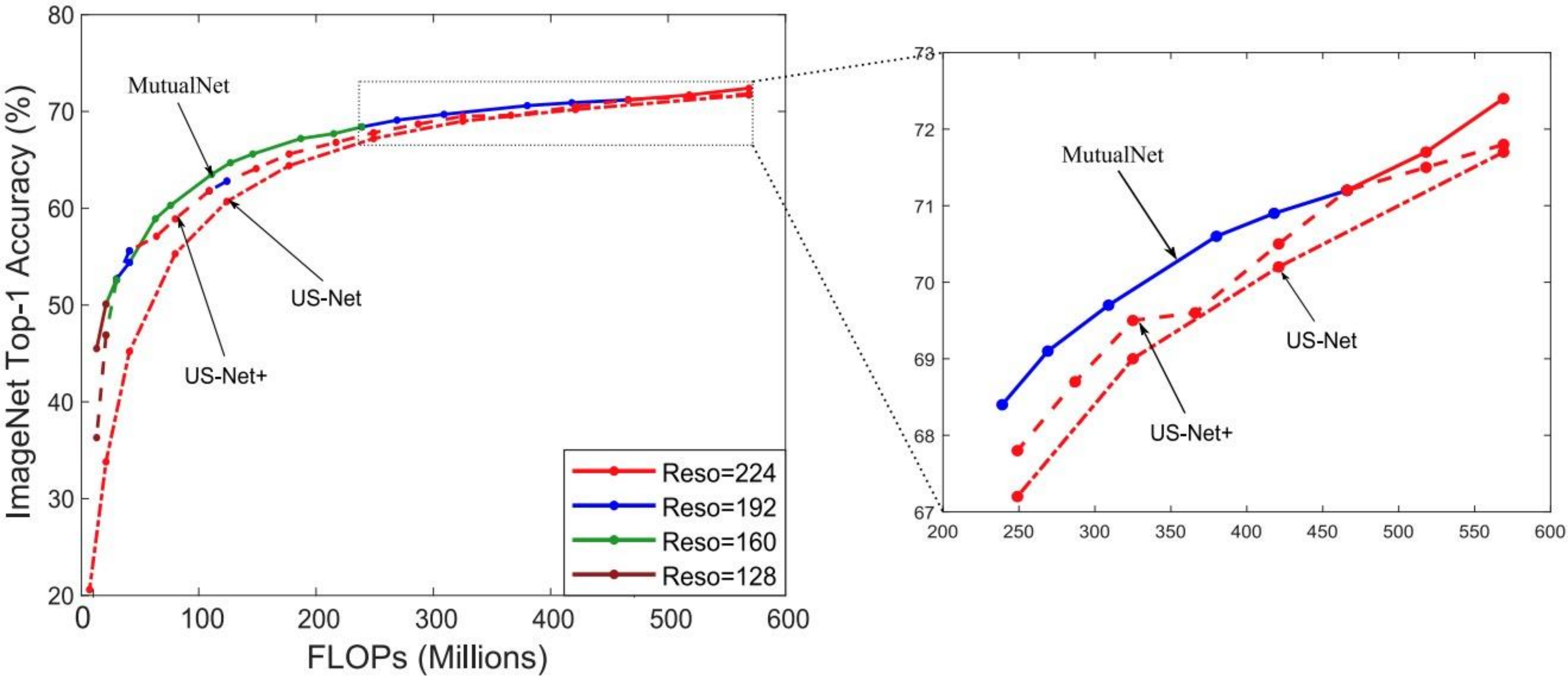


Model	Config	MFLOPs	Top-1 Acc
MobileNet	1.0-224	300	71.8
MutualNet	1.0-224	300	**72.9**
MobileNet	0.75-160	107	66.3
MutualNet	0.75-160	107	**68.3**
MobileNet	0.75-128	69	63.2
MutualNet	0.75-128	69	**64.8**

MobileNet v2

Outperforms individually trained models at different FLOPs constraints

Analysis

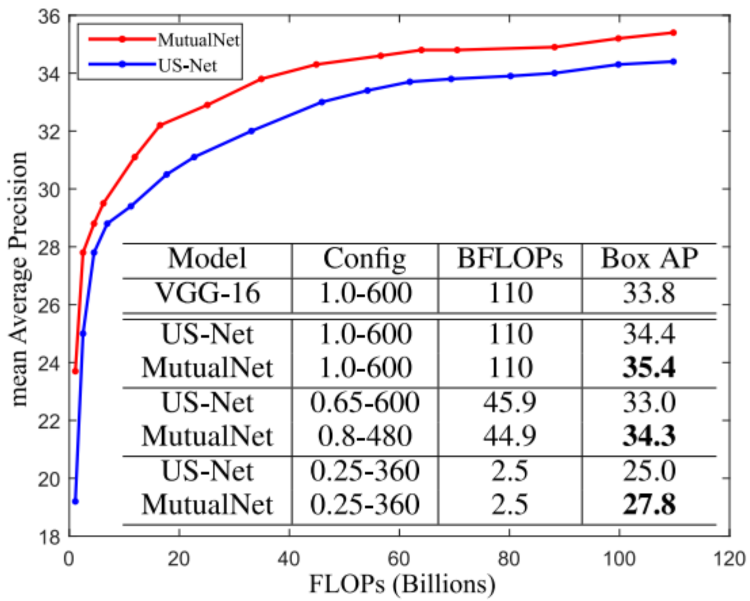


The Accuracy-FLOPs curves are based on MobileNet v1 backbone

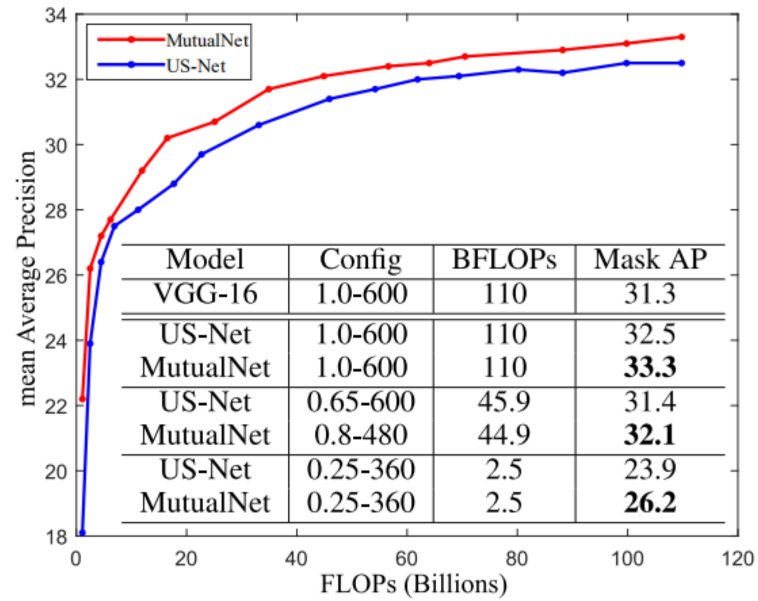
Results – Detection and Segmentation

COCO Dataset: 118K training images and 5K validation images in 80 object categories.

COCO object detection and instance segmentation



Model	Config	BFLOPs	Box AP
VGG-16	1.0-600	110	33.8
US-Net	1.0-600	110	34.4
MutualNet	1.0-600	110	**35.4**
US-Net	0.65-600	45.9	33.0
MutualNet	0.8-480	44.9	**34.3**
US-Net	0.25-360	2.5	25.0
MutualNet	0.25-360	2.5	**27.8**



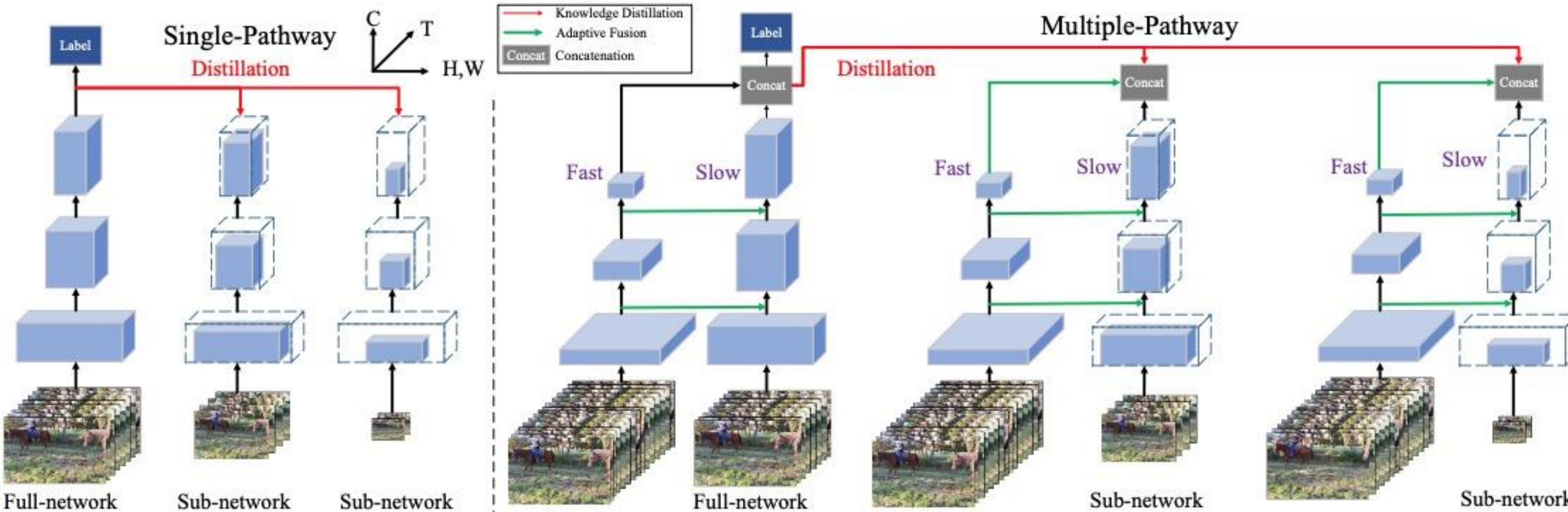
Model	Config	BFLOPs	Mask AP
VGG-16	1.0-600	110	31.3
US-Net	1.0-600	110	32.5
MutualNet	1.0-600	110	**33.3**
US-Net	0.65-600	45.9	31.4
MutualNet	0.8-480	44.9	**32.1**
US-Net	0.25-360	2.5	23.9
MutualNet	0.25-360	2.5	**26.2**

Results – Detection and Segmentation

Code is available at https://github.com/taoyang1122/MutualNet

Dynamic Networks – Research Extensions

- Spatial-temporal domain (e.g., video action recognition)

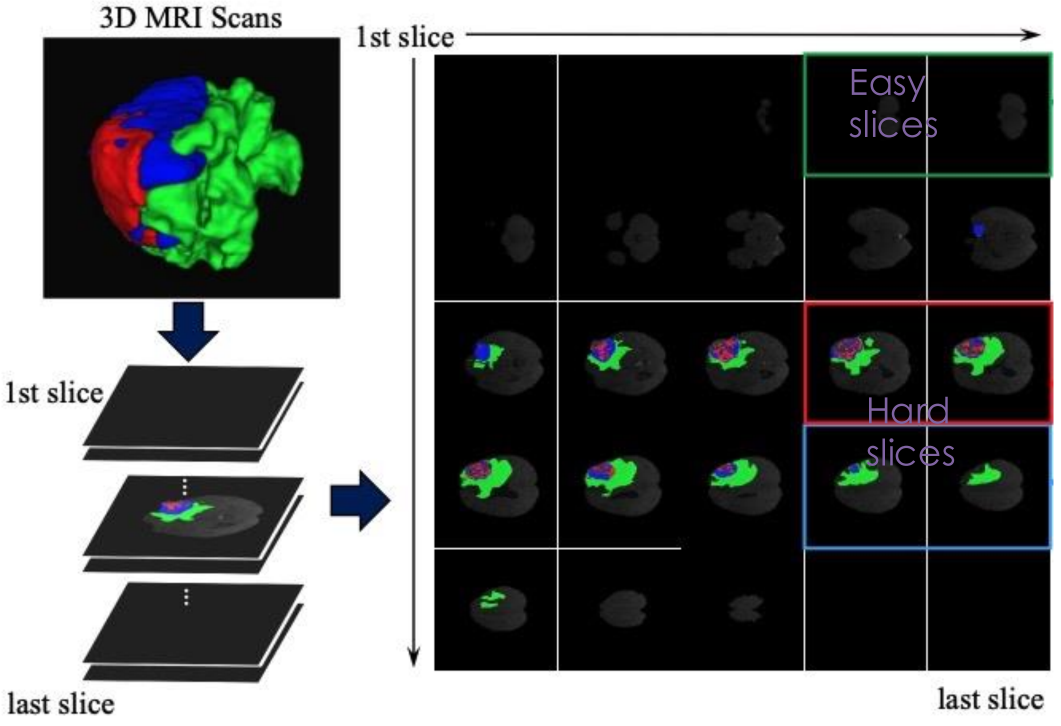


Yang, Taojiannan, Sijie Zhu, Matias Mendieta, Pu Wang, Ravikumar Balakrishnan, Minwoo Lee, Tao Han, Mubarak Shah, and Chen Chen. "MutualNet: Adaptive convnet via mutual learning from different model configurations." IEEE Transactions on Pattern Analysis and Machine Intelligence 45, no. 1 (2023): 811-827.

Med-DANet: Dynamic Architecture Network for Efficient Medical Volumetric Segmentation

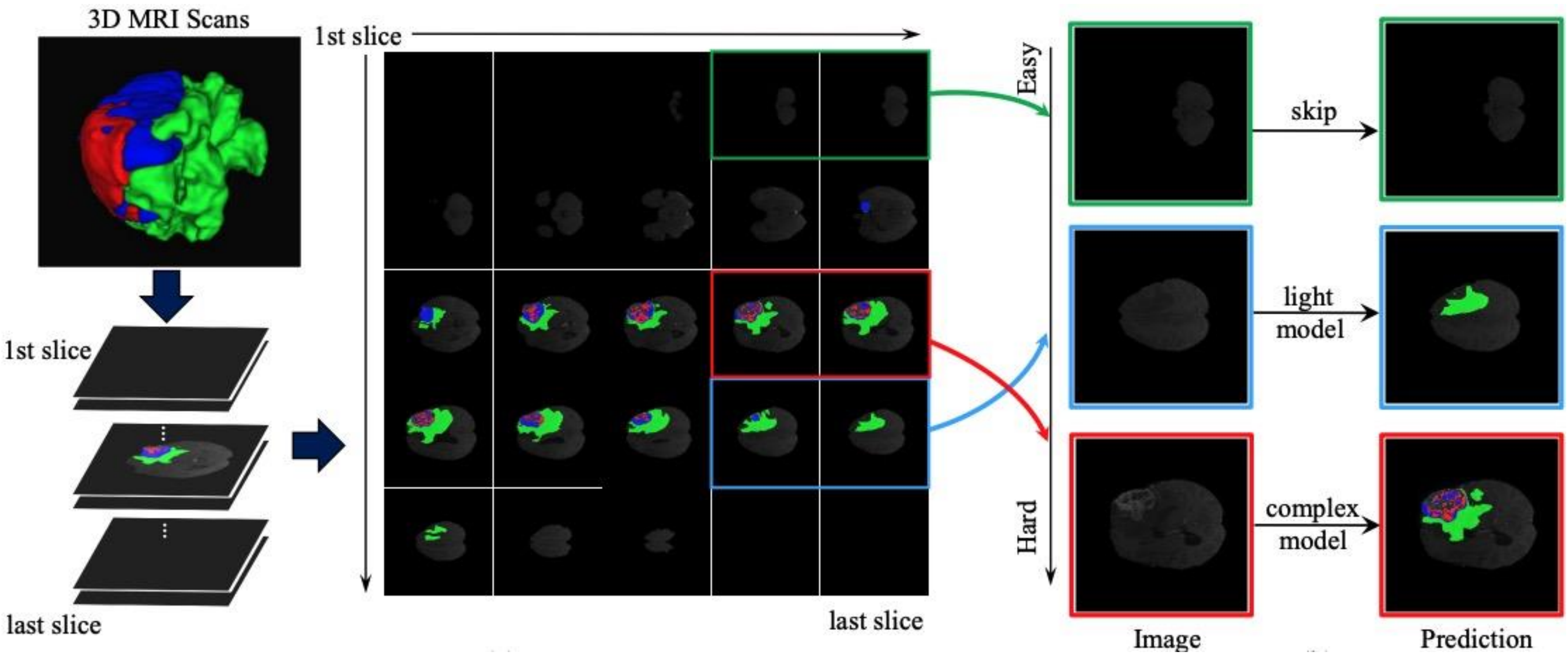
Wang, Wenxuan, Chen Chen, Jing Wang, Sen Zha, Yan Zhang, and Jiangyun Li. "Med-DANet: Dynamic Architecture Network for Efficient Medical Volumetric Segmentation." In European Conference on Computer Vision (ECCV), pp. 506-522. Cham: Springer Nature Switzerland, 2022.

Motivation



Is it necessary to run the same (heavy) model on all the slices to achieve good segmentation results?

Motivation

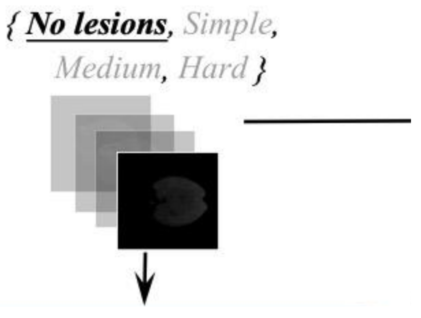


3D MRI Scans
1st slice
last slice
1st slice
last slice
Easy
Hard
skip
light model
complex model
Image
Prediction

Research Question

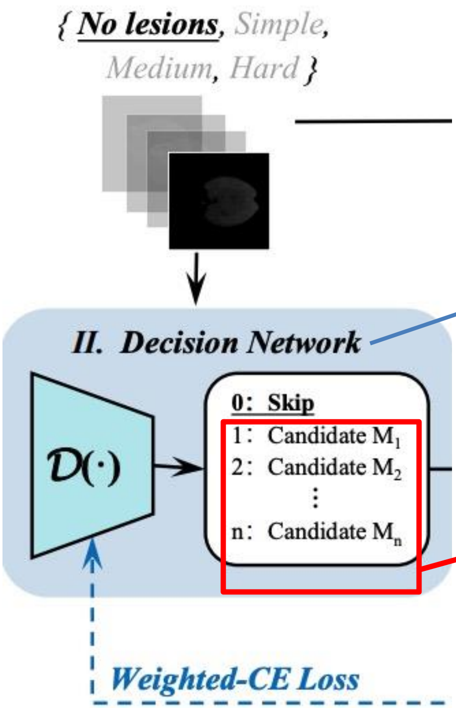
- Is it possible to achieve dynamic inference with adjustable network structures for **better accuracy and efficiency trade-offs** by considering the characteristics of the input data (e.g., the level of segmentation difficulty of each image slice)?

Method - MedDANet

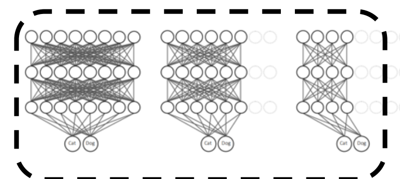


→ Train & Test ← - Train Decision Network only ← - Train Model Bank only Performance Evaluation

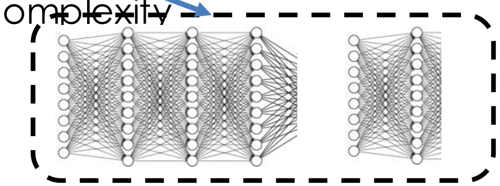
Method - MedDANet



The decision network or policy network is a very lightweight network.

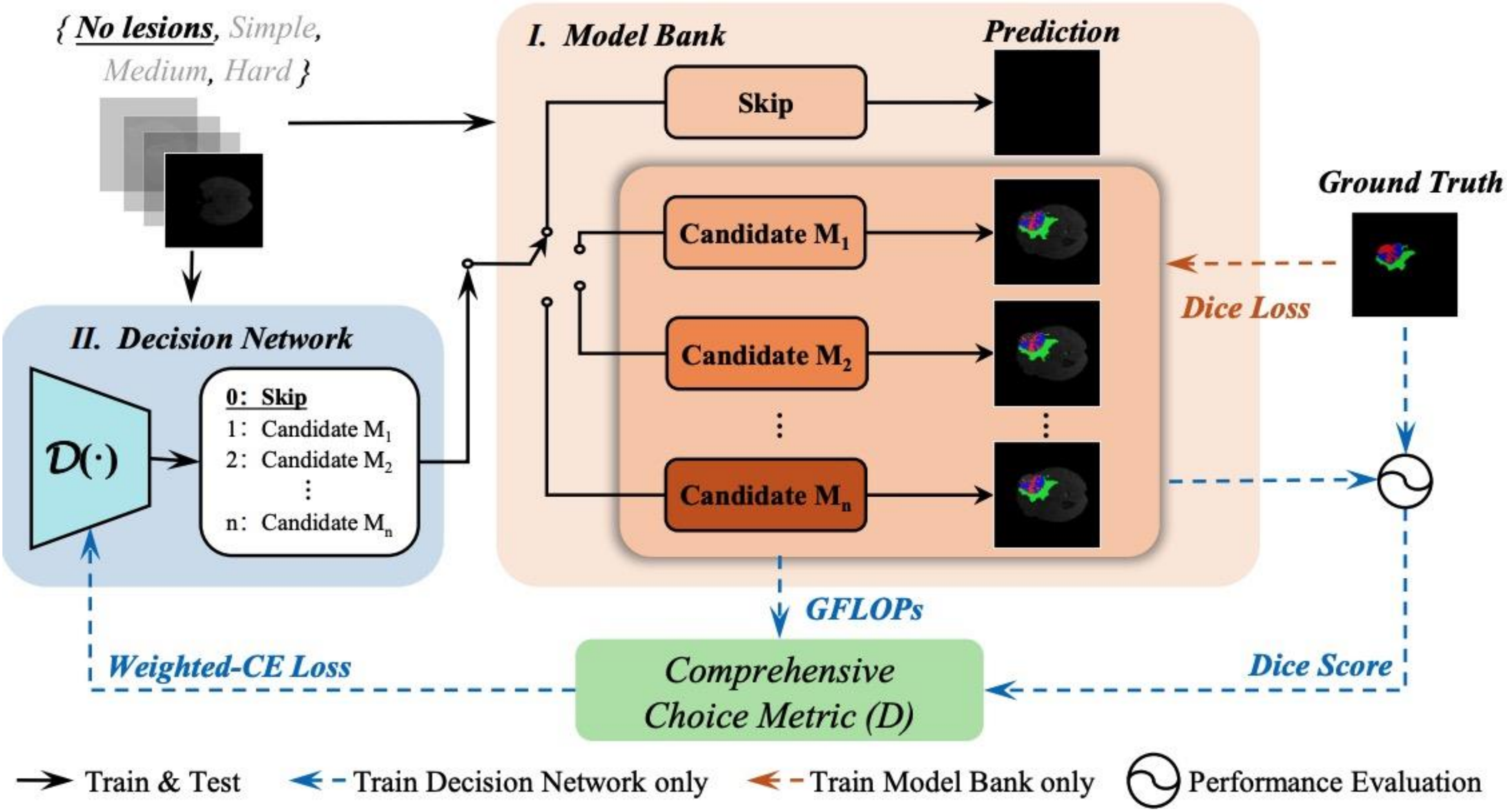


2D UNet with various channel sizes (i.e., model width) or with different scales (i.e., model depth), leading to models with different complexity



Train & Test | Train Decision Network only | Train Model Bank only | Performance Evaluation

Method - MedDANet



Method - MedDANet

Decision network training

- We reduce the channel size of ShuffleNetV2 [24] to get an extremely lightweight classification network as our Decision Network so that its computational overhead is negligible in the entire framework.

$$D = \begin{cases} 0, & P_f < 1 \\ argmax((1-\alpha) * S_i + \alpha * softmax(\frac{1}{F_i})) + 1, & P_f \geqslant 1 \end{cases}, \quad (4)$$

where S_i and F_i is respectively the Dice Score and FLOPs of candidate model M_i during the model training. P_f denotes the number of foreground pixels (all pixels of segmentation targets). Specifically, if the number of foreground pixels is less than 1 (i.e. $P_f < 1$), the current slice will be considered without any lesion areas, which should be directly skipped (i.e. the corresponding supervision is 0)

Results

Table 1. Performance comparison on BraTS 2019 validation set.

Method	Dice Score (%) ↑			Hausdorff Dist. (mm) ↓			FLOPs (G) ↓	
	ET	WT	TC	ET	WT	TC	per case	per slice
3D U-Net [11]	70.86	87.38	72.48	5.062	9.432	8.719	1,669.53	13.04
V-Net [27]	73.89	88.73	76.56	6.131	6.256	8.705	749.29	5.85
Attention U-Net [29]	75.96	88.81	77.20	5.202	7.756	8.258	132.67	1.04
Wang et al. [35]	73.70	89.40	80.70	5.994	5.677	7.357	-	-
Chen et al. [8]	74.16	**90.26**	79.25	4.575	**4.378**	7.954	-	-
Li et al. [18]	77.10	88.60	81.30	6.033	6.232	7.409	-	-
Frey et al. [12]	78.70	89.60	80.00	6.005	8.171	8.241	-	-
TransUNet [7]	78.17	89.48	78.91	4.832	6.667	7.365	1205.76	9.42
Swin-UNet [4]	78.49	89.38	78.75	6.925	7.505	9.260	250.88	1.96
TransBTS [36]	78.36	88.89	**81.41**	5.908	7.599	7.584	333.09	2.60
Ours	**79.99**	90.13	80.83	**4.086**	5.826	**6.886**	**77.78**	**0.61**

Results

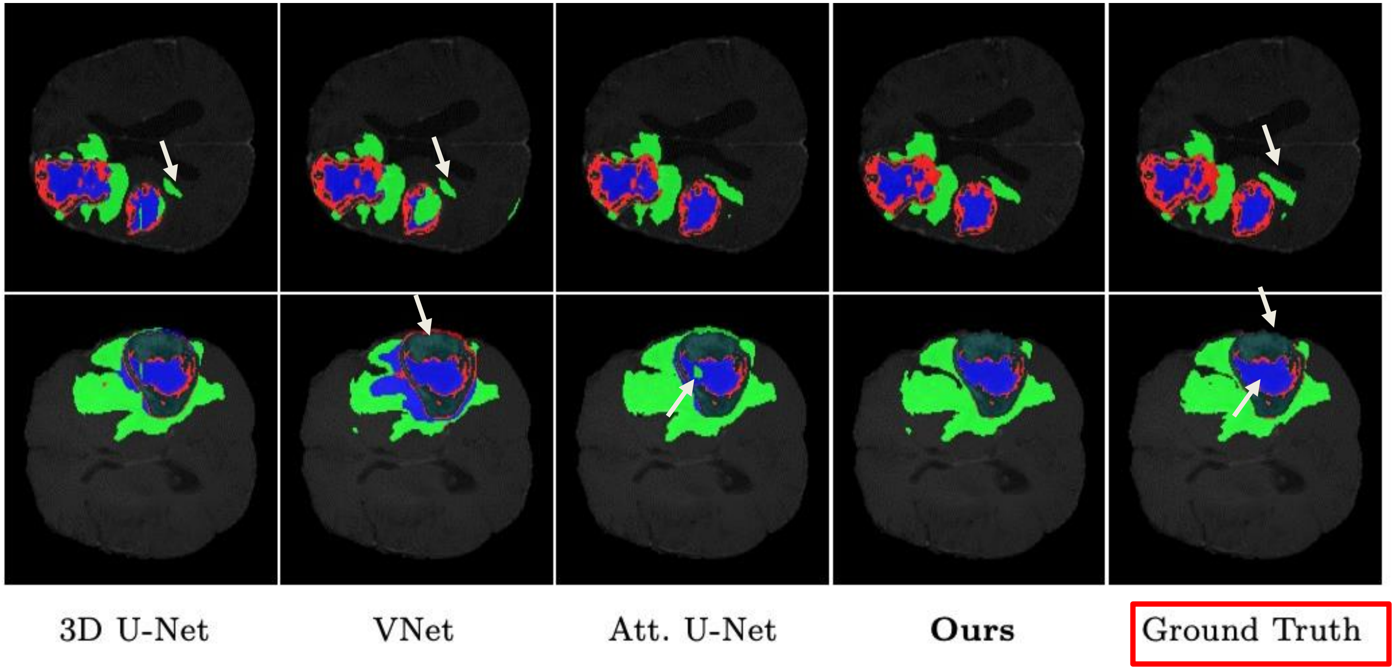


Fig. 3. The visual comparison of MRI brain tumor segmentation results. The blue regions denote the enhancing tumors, the red regions denote the non-enhancing tumors, and the green ones denote the peritumoral edema.

Results

Liver Tumor Segmentation using CT scans

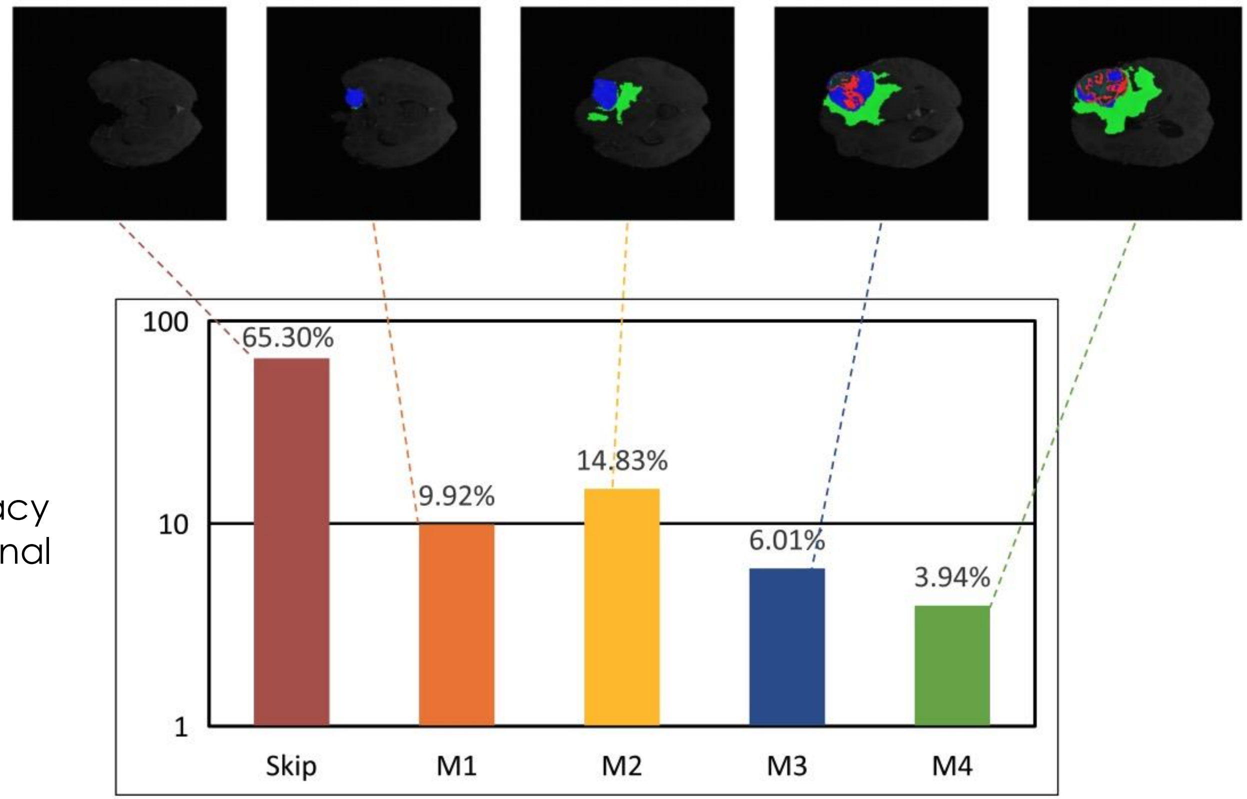
Table 7. Performance comparison on LiTS 2017 testing set. "P" refers to pre-trained model. Per case and per slice denote the computational cost of segmenting a 3D patient case and a single 2D slice, respectively.

Method	Dice per case (%) ↑		Dice global (%) ↑		FLOPs (G) ↓	
	Lesion	Liver	Lesion	Liver	Per Case	Per Slice
U-Net [10]	65.00	-	-	-	-	-
3D DenseUNet w/o P [19]	59.40	93.60	78.80	92.90	-	-
2D DenseUNet w/o P [19]	67.70	94.70	80.10	94.70	-	-
2D DenseNet w/ P [19]	68.30	95.30	81.80	95.90	-	-
2D DenseUNet w/ P [19]	70.20	95.80	**82.10**	96.30	-	-
I3D [5]	62.40	95.70	77.60	96.00	-	-
I3D w/ P [5]	66.60	95.60	79.90	96.20	-	-
Han [14]	67.00	-	-	-	-	-
Vorontsov et al. [33]	65.00	-	-	-	-	-
TransUNet [7]	61.70	95.40	77.40	95.60	1200.64	9.38
Swin-UNet [4]	-	92.70	67.60	91.60	249.60	1.95
TransBTS [36]	70.30	96.00	81.50	96.40	330.00	2.58
Ours	**70.50**	**96.10**	81.90	**96.60**	**37.12**	**0.29**

UCF CENTER FOR RESEARCH IN COMPUTER VISION

Analysis

A good balance between accuracy and computational efficiency



M1 Increasing model complexity M4

Fig. 5. The activation ratio of each candidate model for different medical image slices in BraTS 2019 dataset. Skip, M1, M2, M3, M4 denote the operation of directly skip, candidate 1, candidate 2, candidate 3, and candidate 4, respectively.

- Any ideas on improving the Med-DANet?

Med-DANet-V2

- Shen, Haoran, Yifu Zhang, Wenxuan Wang, Chen Chen, Jing Liu, Shanshan Song, and Jiangyun Li. "Med-DANet V2: A Flexible Dynamic Architecture for Efficient Medical Volumetric Segmentation." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 7871-7881. 2024.

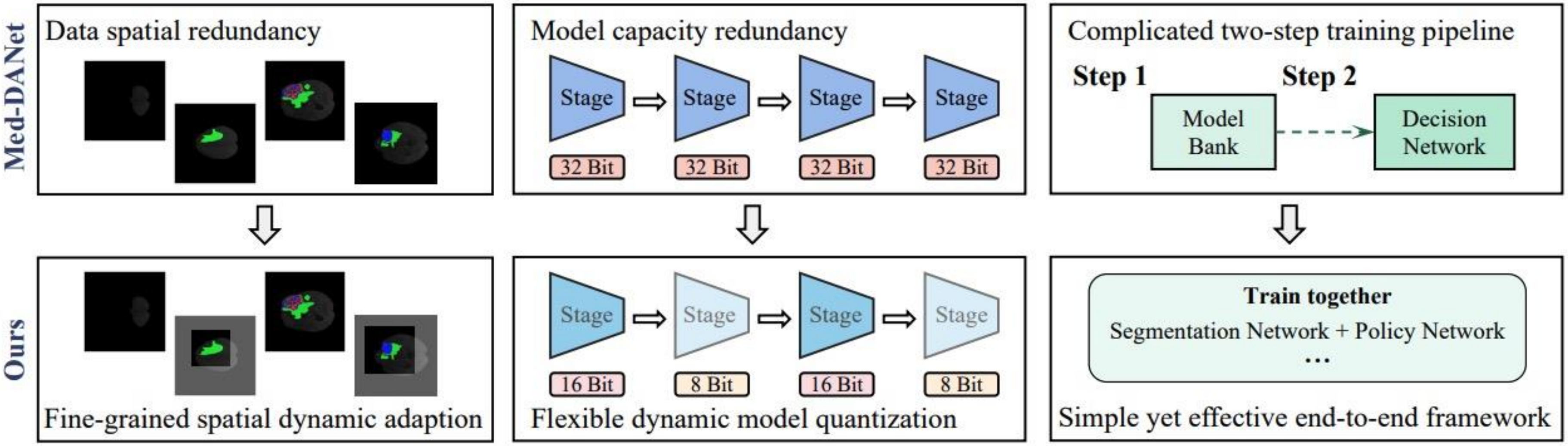
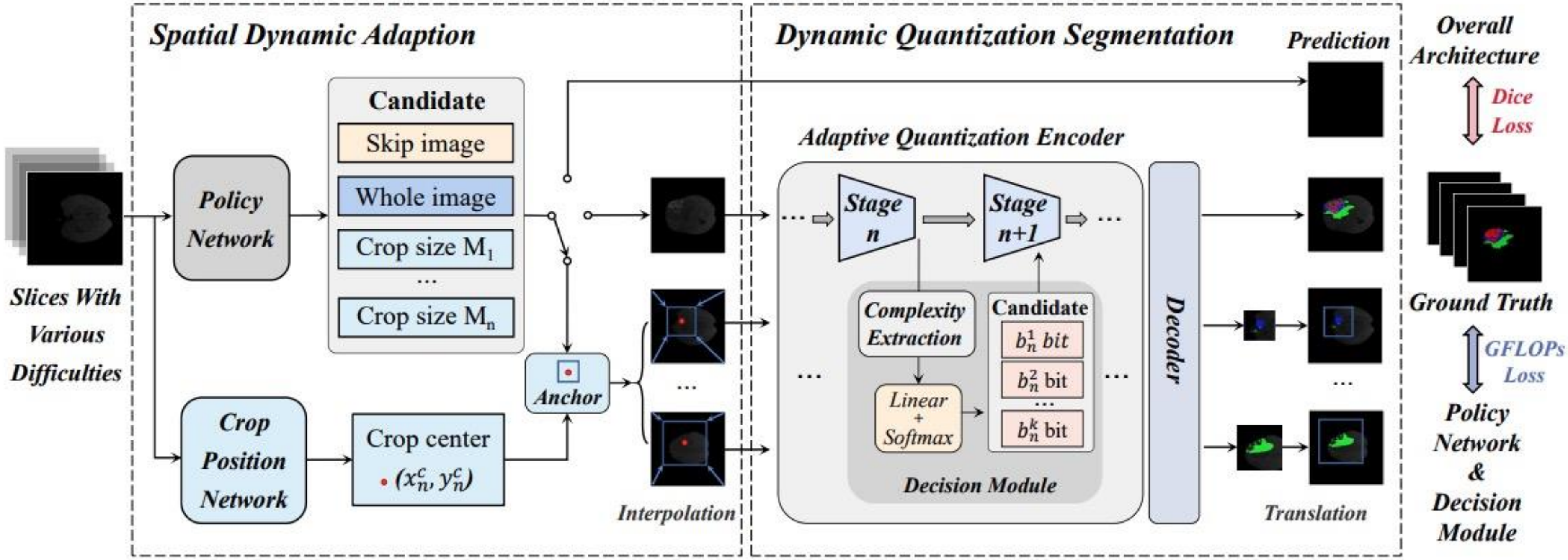


Figure 1. The comparison between the previous dynamic network Med-DANet and our proposed Med-DANet V2 (Ours).

Med-DANet-V2



Spatial Dynamic Adaption
Dynamic Quantization Segmentation
Prediction
Overall Architecture
Dice Loss
Candidate
Skip image
Whole image
Crop size M_1
...
Crop size M_n
Policy Network
Slices With Various Difficulties
Crop Position Network
Crop center
(x_n^c, y_n^c)
Anchor
Interpolation
Adaptive Quantization Encoder
Stage n
Stage n+1
Complexity Extraction
Linear + Softmax
Candidate
b_n^1 bit
b_n^2 bit
b_n^k bit
Decision Module
Decoder
Translation
Ground Truth
GFLOPs Loss
Policy Network & Decision Module

References and resources

1. Blalock, Davis, et al. "What is the state of neural network pruning?." *Proceedings of machine learning and systems* 2 (2020): 129-146.
2. Liang, Tailin, et al. "Pruning and quantization for deep neural network acceleration: A survey." *Neurocomputing* 461 (2021): 370-403.
3. Gholami, Amir, et al. "A survey of quantization methods for efficient neural network inference." arXiv preprint arXiv:2103.13630 (2021).
4. Gou, Jianping, et al. "Knowledge distillation: A survey." *International Journal of Computer Vision* 129.6 (2021): 1789-1819.
5. Wang, Lin, and Kuk-Jin Yoon. "Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
6. https://github.com/lilujunai/Awesome-Knowledge-Distillation-for-CV

Thank you!

Question?