Dynamic Neural Networks: A Survey

Yizeng Han*, Gao Huang*, Shiji Song, Le Yang, Honghui Wang, Yulin Wang

Department of Automation, Tsinghua University

清華大學 Tsinghua University

Overview

清華大學
Tsinghua University

- Introduction
- Sample-wise Dynamic Networks
- Spatial-wise Dynamic Networks
- Temporal-wise Dynamic Networks
- Inference & Training
- Applications
- Discussion

Overview

Tsinghua University

- **Introduction**
- Sample-wise Dynamic Networks
- Spatial-wise Dynamic Networks
- Temporal-wise Dynamic Networks
- Inference & Training
- Applications
- Discussion

Why architecture matters?

Representation power

Optimization Characteristics

Generalization

Efficiency

Advances in Neural Architecture Design

Tsinghua University

- **AlexNet**
- ZF-Net
- DSN
- NIN
- **VGG**
- **GoogleNet**
- …

2012-2015

Fast developing stage

High diversity

- Highway Networks
- FractalNet
- **ResNet**
- **DenseNet**
- ResNeXt
- Dual Path Network
- …

2015-2017

Mature stage

Starting to converge

2017-Present

Prosperous stage

Very high diversity

Light-weighted models

- **MobileNet** (V1, V2, V3)
- **CondenseNet** (V1, V2)
- **ShuffleNet** (V1, V2)
- …

Neural Arch. Search

- **NASNet**
- DARTS
- …

Dynamic models

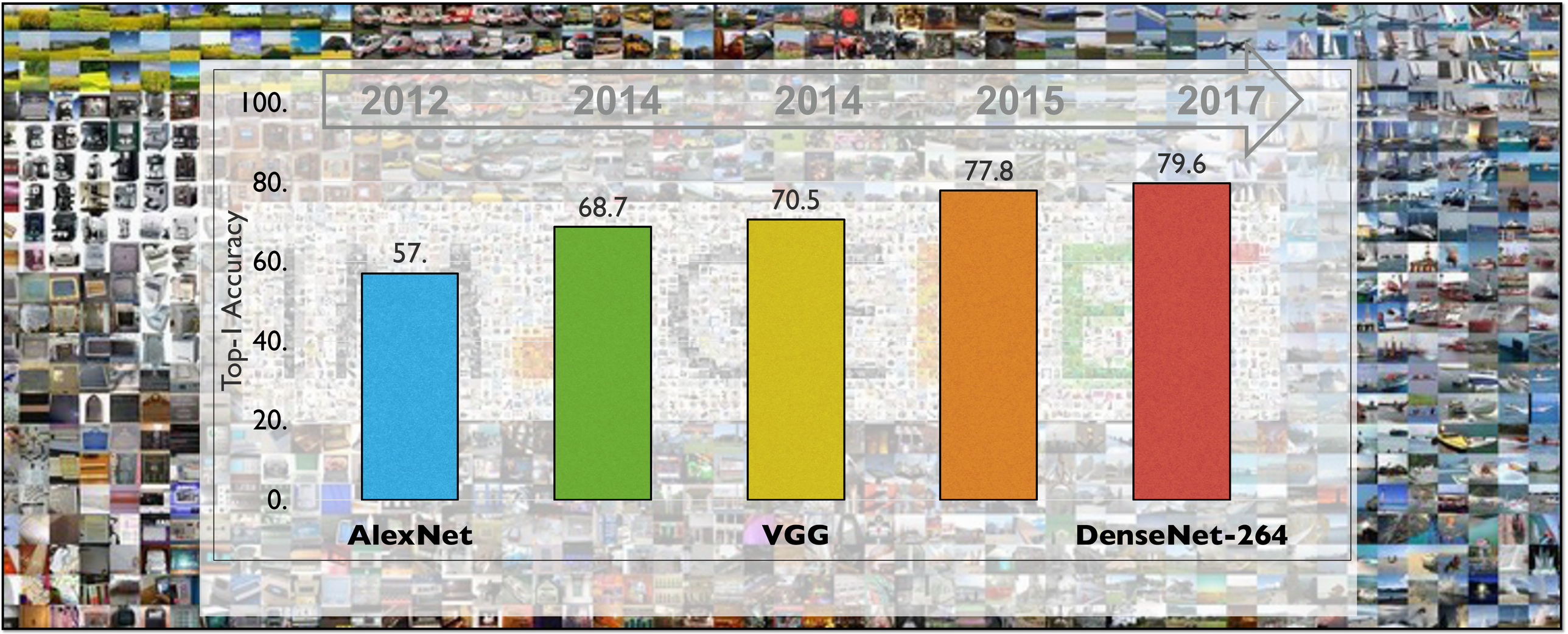
- **MSDNet**
- Block-Drop
- Glance and Focus
- …

Transformer?!

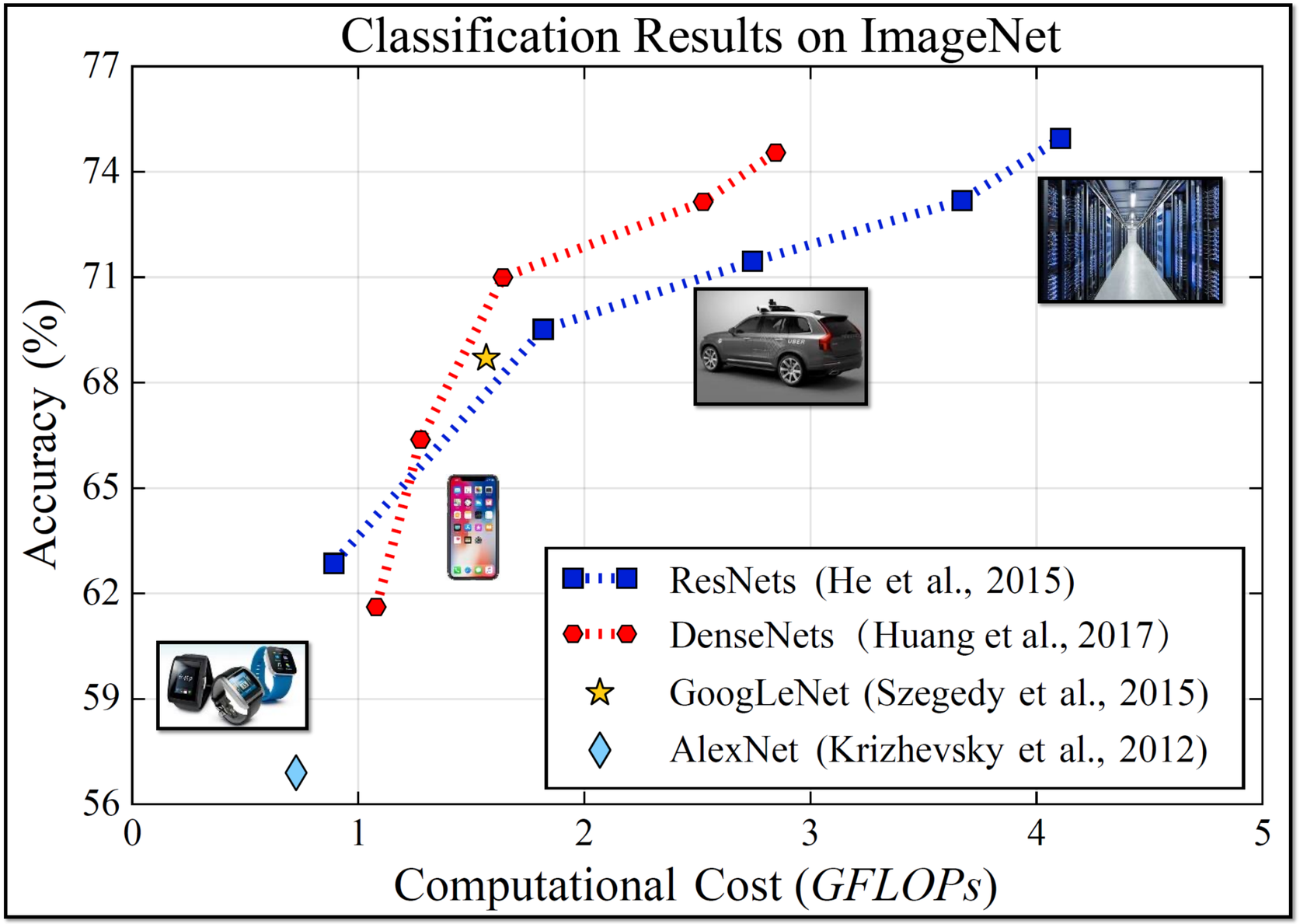
- …

Dynamic Neural Networks

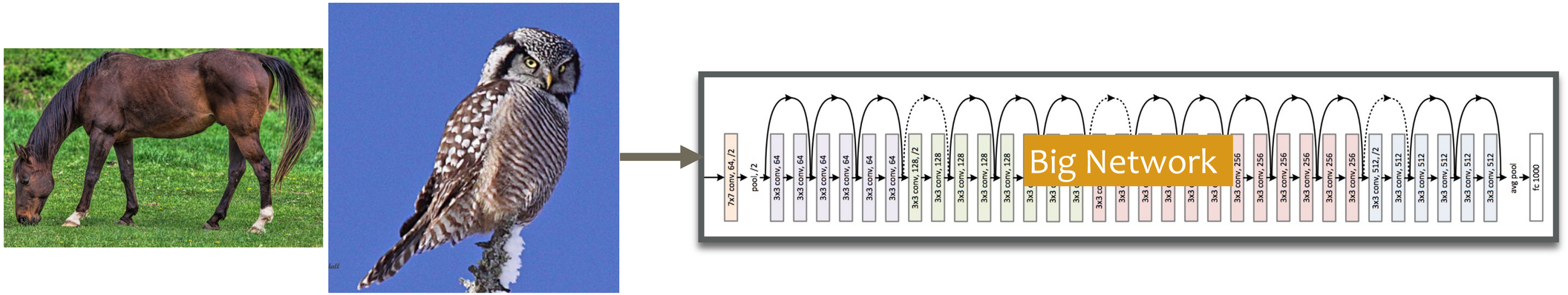
Development of Deep Learning



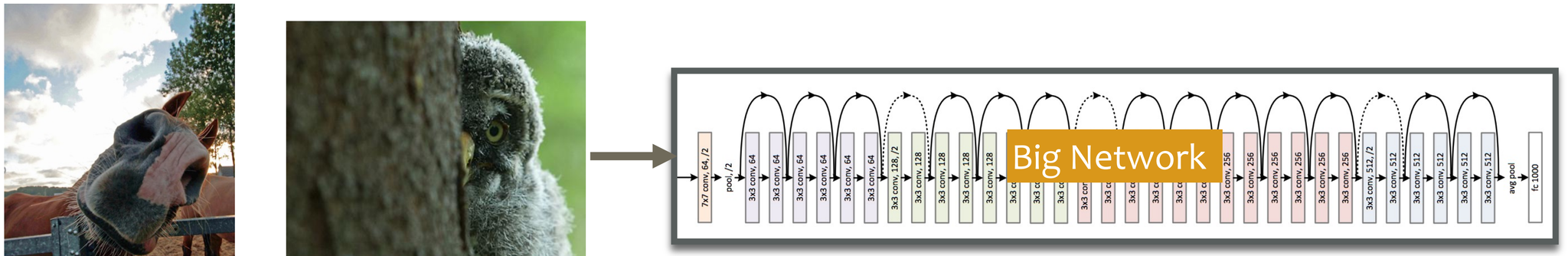
Accuracy-Time Tradeoff



*Most conventional neural networks recognize all instances with **the same architecture**.*

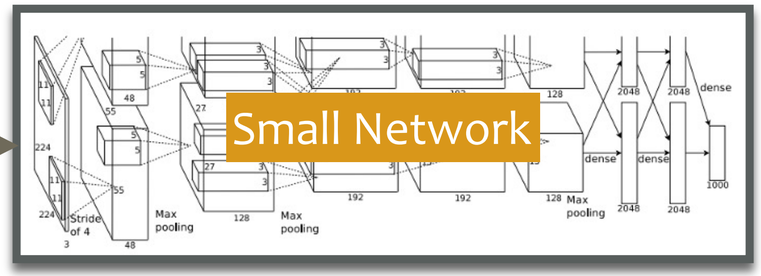


Canonical (“easy”)

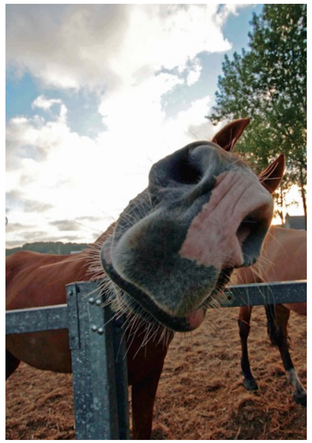


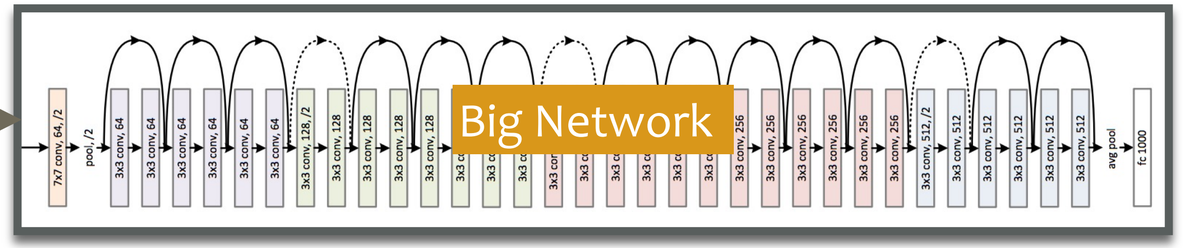
Noncanonical (“hard”)

A naïve idea of adaptive inference



Canonical (“easy”)





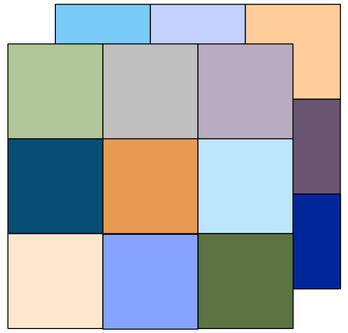
Noncanonical (“hard”)

Dynamic networks can
***adapt** their **architectures** to each sample.*

*Most conventional neural networks recognize all instances with **the same parameters**.*

Input

*

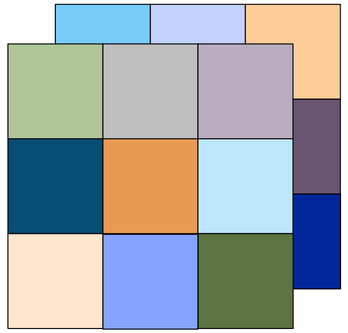


Conv Kernel

*Most conventional neural networks recognize all instances with **the same parameters**.*

Input

*



Conv Kernel

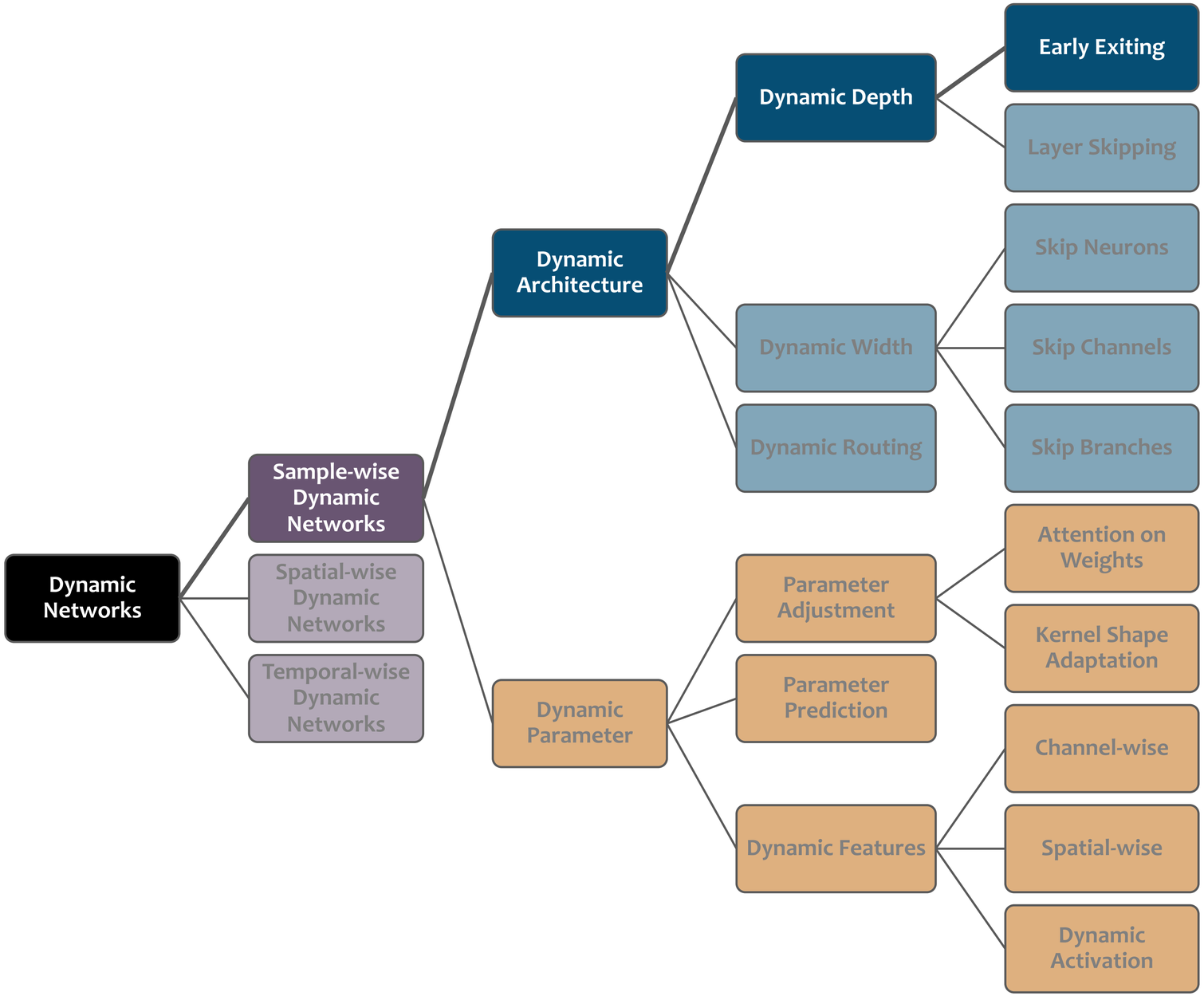
Dynamic networks can
***adapt** their **parameters** to each sample.*

Overview

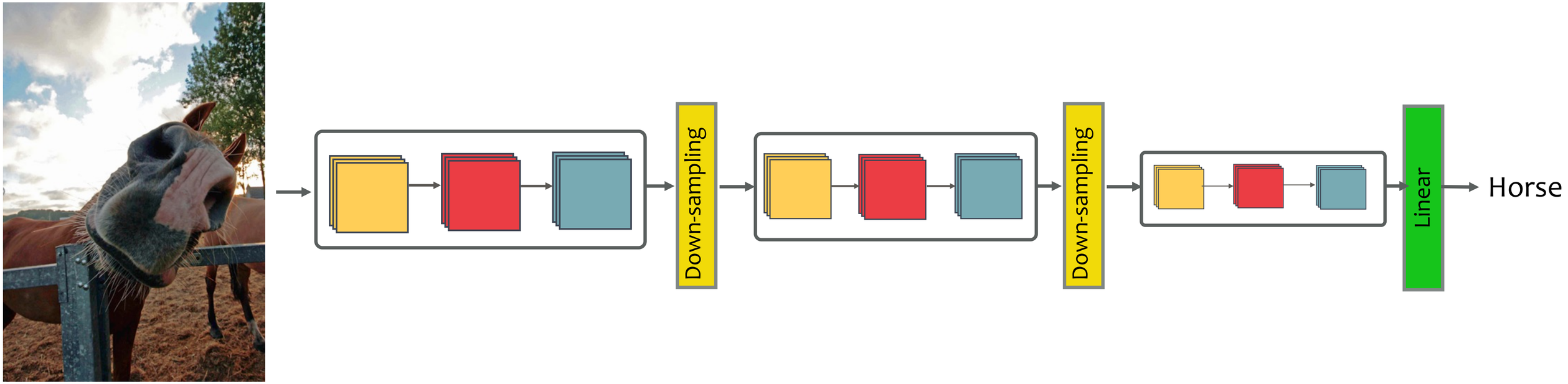
清華大學
Tsinghua University

- Introduction
- **Sample-wise Dynamic Networks**
- Spatial-wise Dynamic Networks
- Temporal-wise Dynamic Networks
- Inference & Training
- Applications
- Discussion

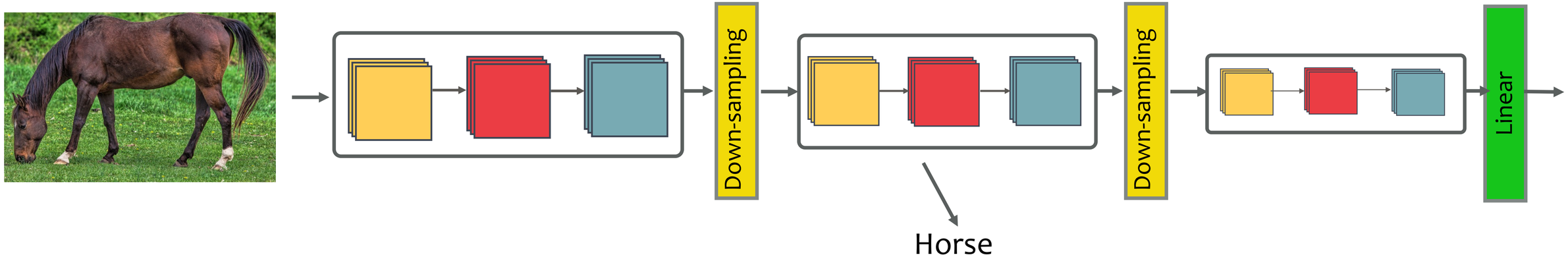
Sample-wise Dynamic Neural Networks



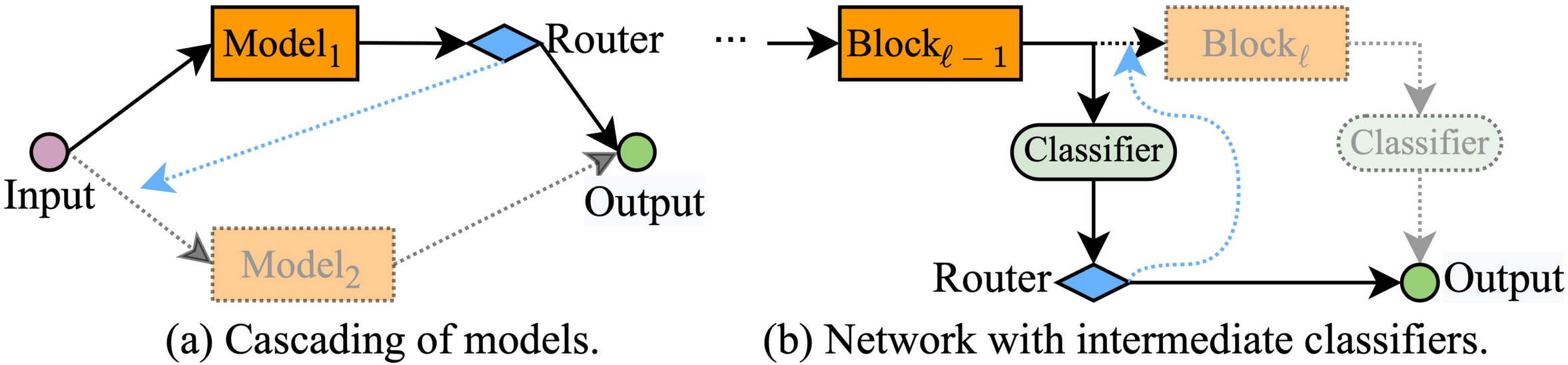
Early Exiting



Early Exiting



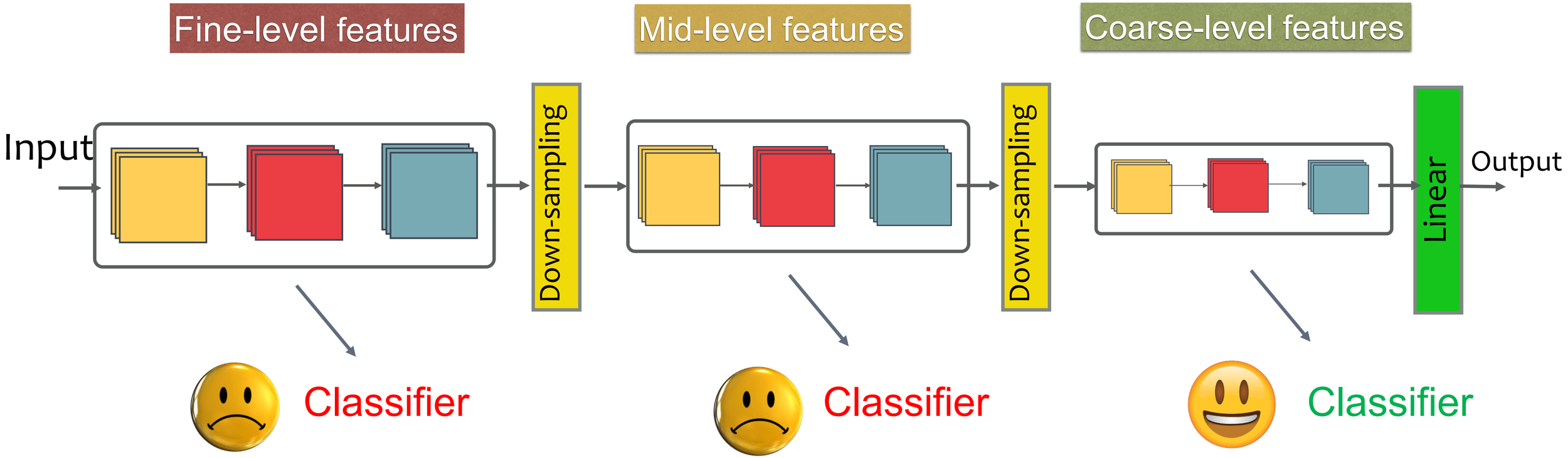
Early Exiting: Two Implementations



- Park, E., Kim, D., Kim, S., Kim, Y. D., Kim, G., Yoon, S., & Yoo, S. (2015, October). Big/little deep neural network for ultra low power inference. In *2015 International Conference on Hardware/Software Codesign and System Synthesis (CODES+ ISSS)* (pp. 124-132). IEEE.
- Teerapittayanon, S., McDanel, B., & Kung, H. T. (2016, December). Branchynet: Fast inference via early exiting from deep neural networks. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 2464-2469). IEEE.
- Bolukbasi, T., Wang, J., Dekel, O., & Saligrama, V. (2017, July). Adaptive neural networks for efficient inference. In International Conference on Machine Learning (pp. 527-536). PMLR.

Early Exiting

A challenge: Intermediate classifiers may interfere with each other



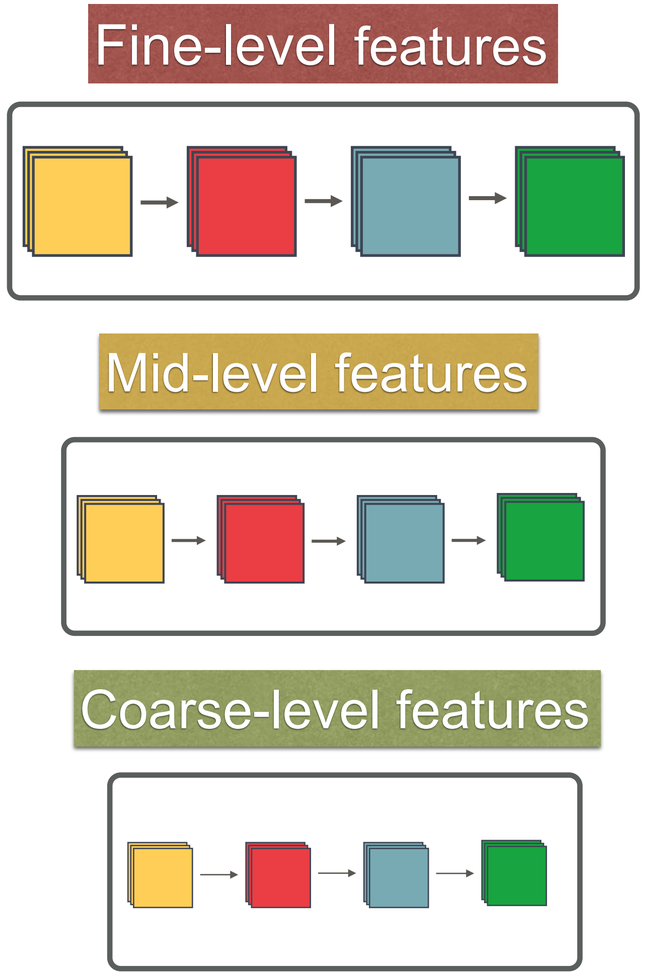
Classifiers **only work well** on **coarse-scale** feature maps

Nearly **all computation** has been done **before** getting a **coarse** feature

Dynamic Depth: Early Exiting

Tsinghua University

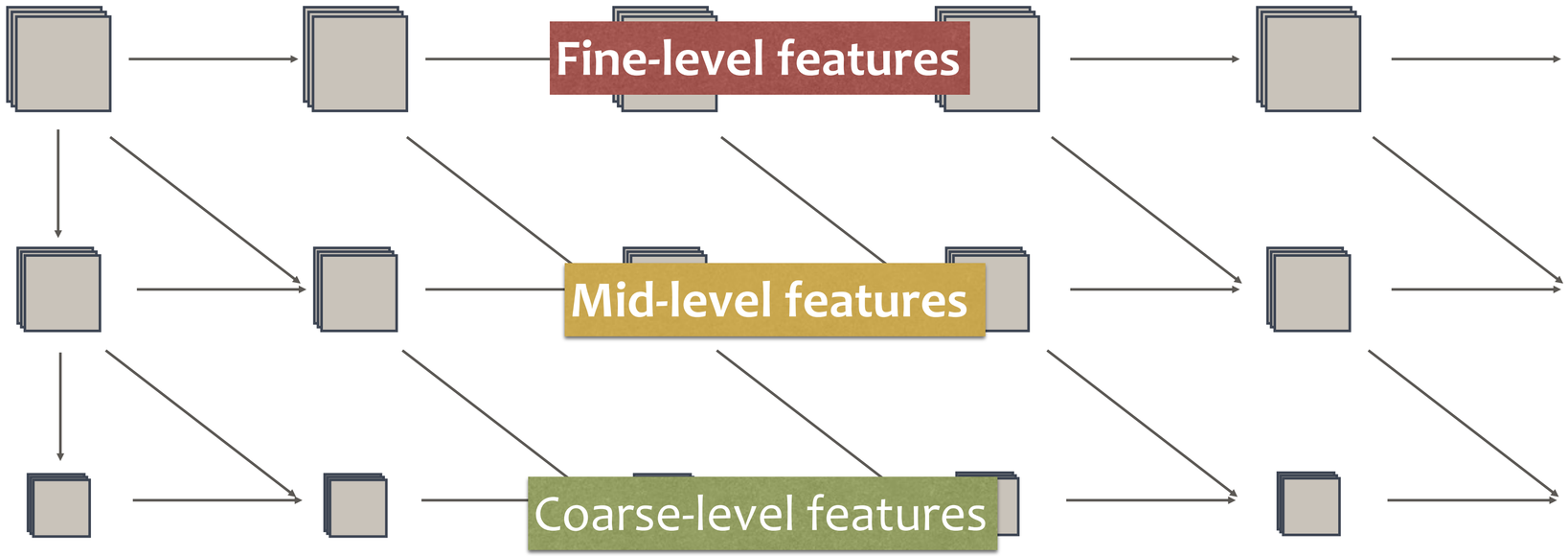
Solution: Multi-scale Architecture



Dynamic Depth: Early Exiting

Tsinghua University

Solution: Multi-scale Architecture

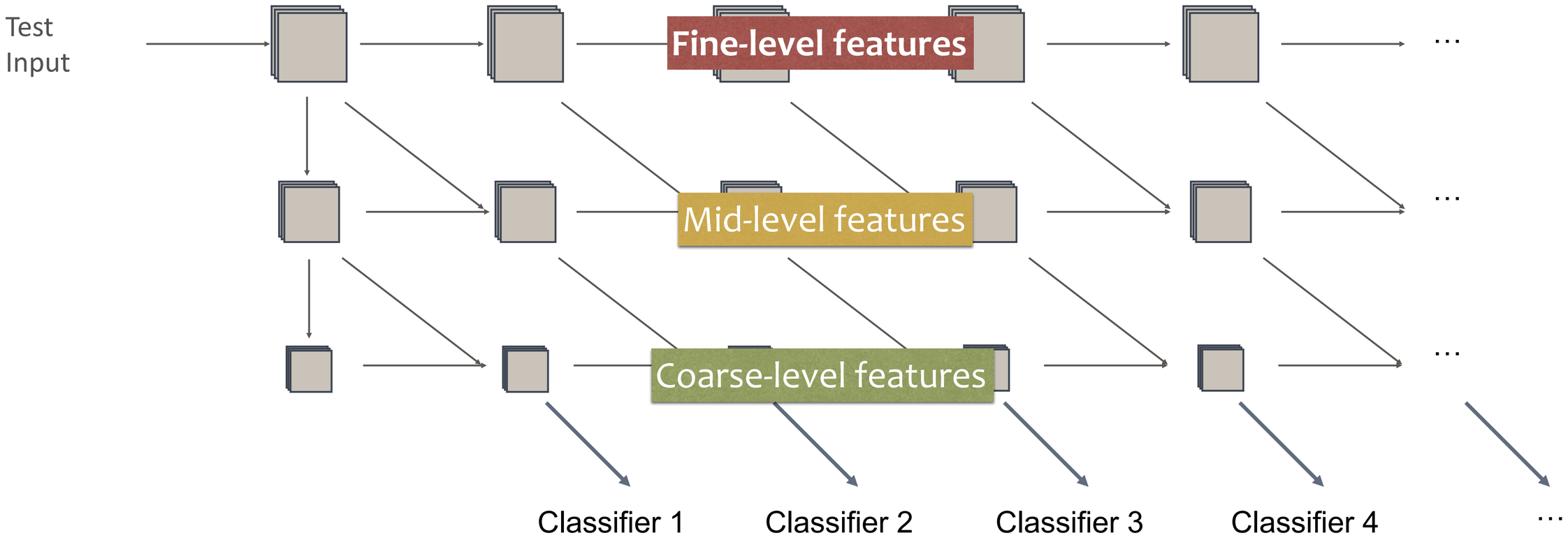


- Huang, G., Chen, D., Li, T., Wu, F., van der Maaten, L., & Weinberger, K. (2018, February). Multi-Scale Dense Networks for Resource Efficient Image Classification. In International Conference on Learning Representations.

Dynamic Depth: Early Exiting

Tsinghua University

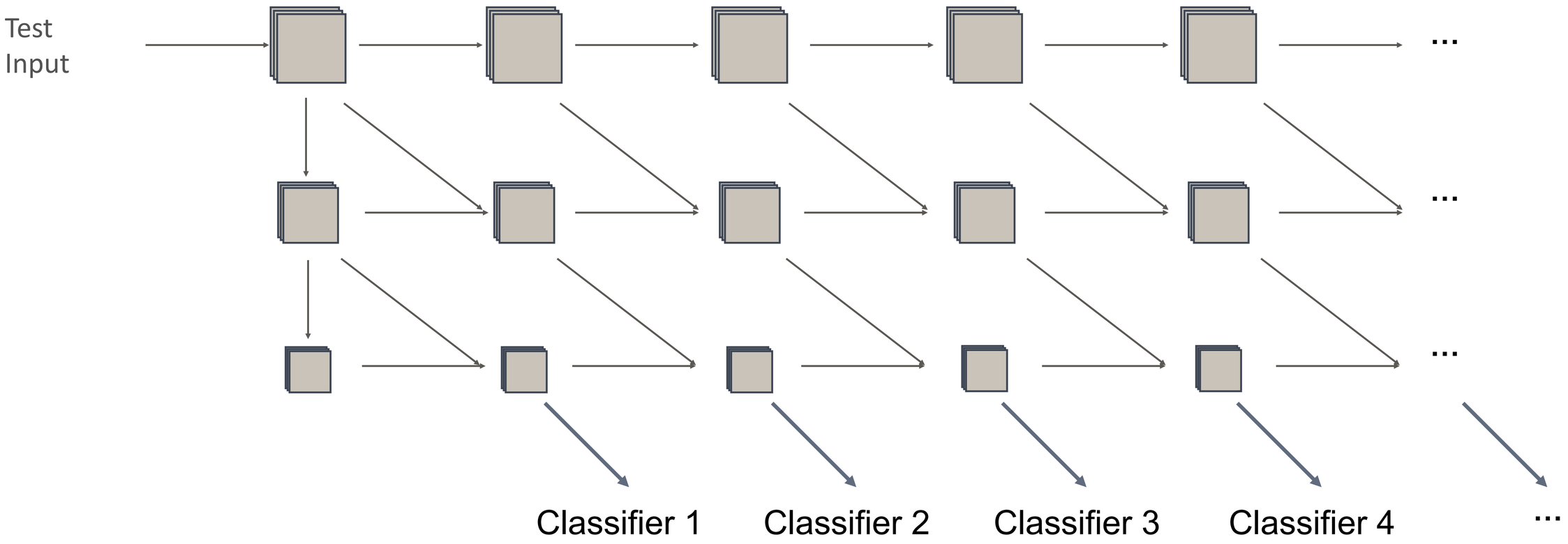
Solution: Multi-scale Architecture



Classifiers only operate on high level features!

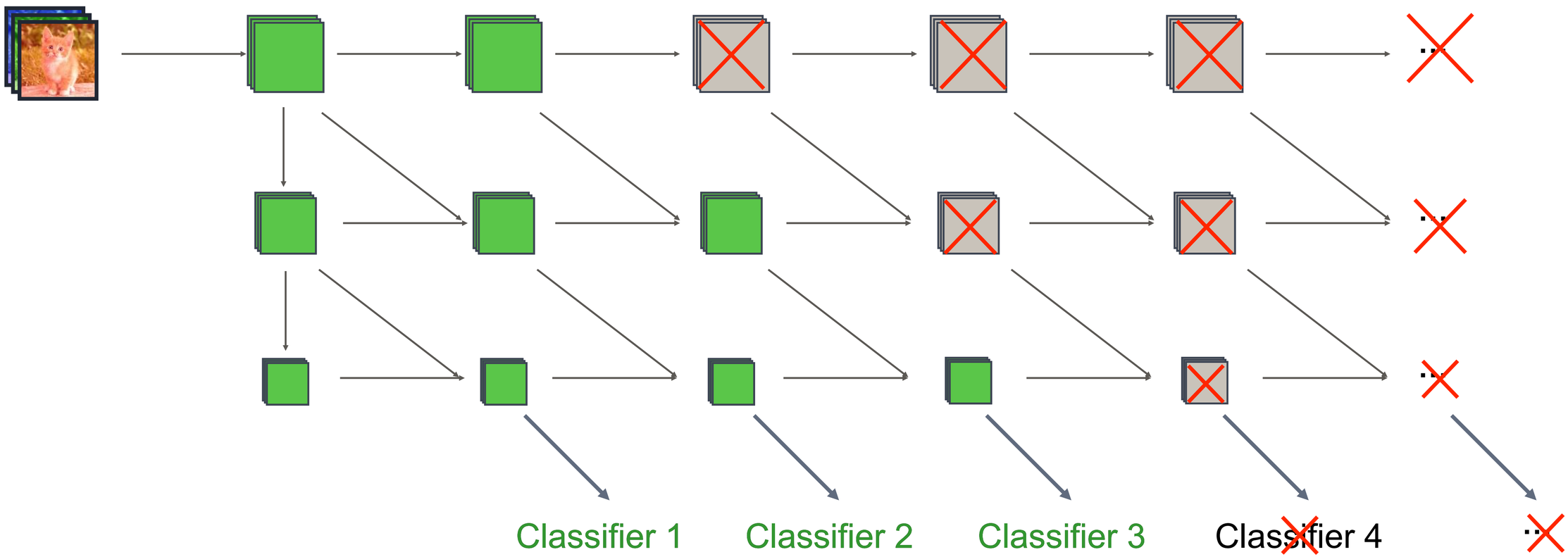
- Huang, G., Chen, D., Li, T., Wu, F., van der Maaten, L., & Weinberger, K. (2018, February). Multi-Scale Dense Networks for Resource Efficient Image Classification. In International Conference on Learning Representations.

Multi-scale densenet



- Huang, G., Chen, D., Li, T., Wu, F., van der Maaten, L., & Weinberger, K. (2018, February). Multi-Scale Dense Networks for Resource Efficient Image Classification. In International Conference on Learning Representations.

Multi-scale densenet



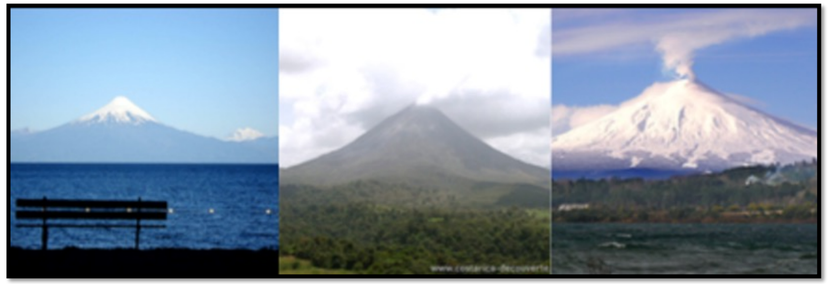
- Huang, G., Chen, D., Li, T., Wu, F., van der Maaten, L., & Weinberger, K. (2018, February). Multi-Scale Dense Networks for Resource Efficient Image Classification. In International Conference on Learning Representations.

Dynamic Depth: Early Exiting

Tsinghua University

Class:
red wine

Class:
volcano

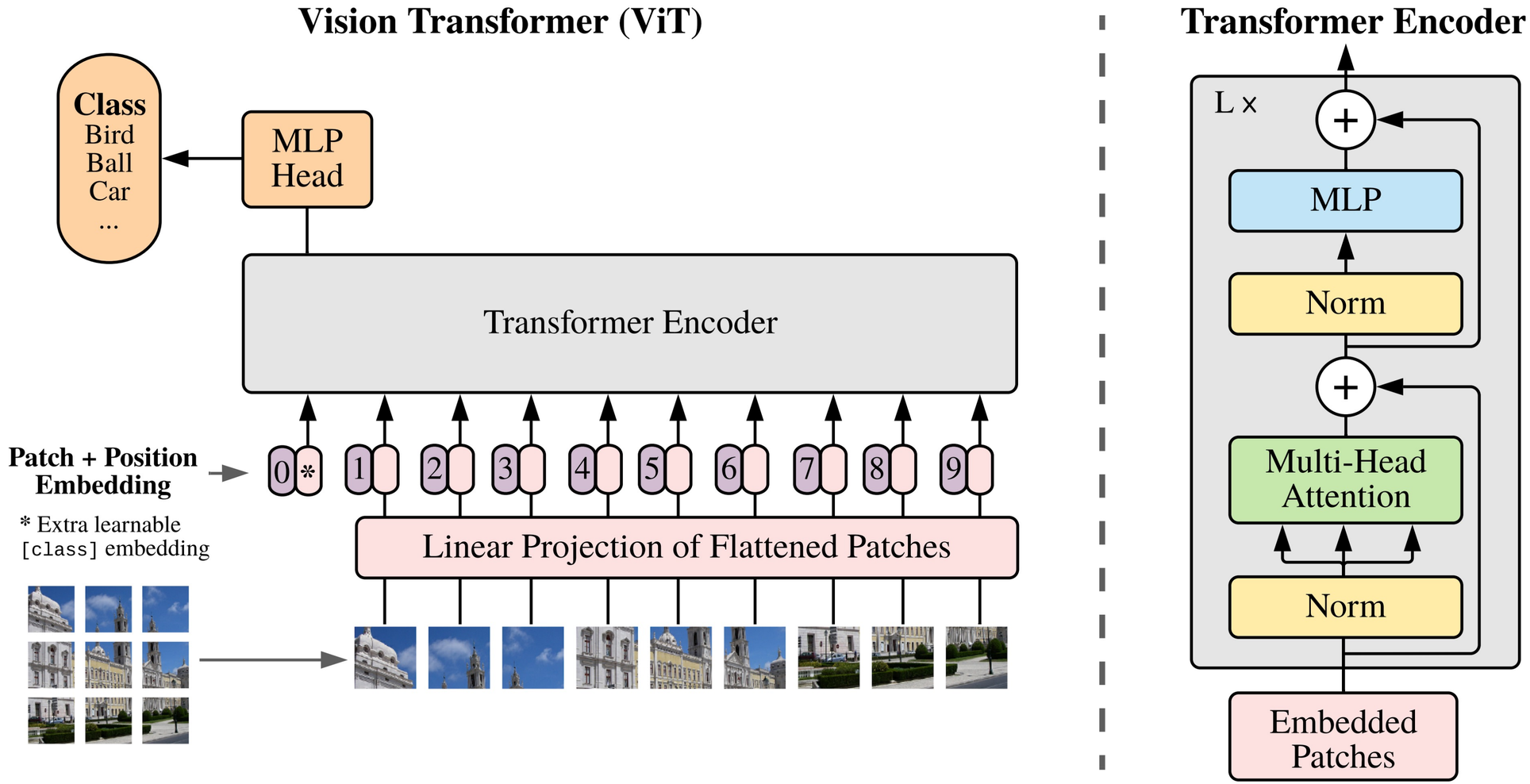


"easy"
(exit at **first** classifier)

"hard"
(exit at **last** classifier)

- Huang, G., Chen, D., Li, T., Wu, F., van der Maaten, L., & Weinberger, K. (2018, February). Multi-Scale Dense Networks for Resource Efficient Image Classification. In International Conference on Learning Representations.

Dynamic Vision Transformers



The ViT structure

- Dosovitskiy et al. (2020, Nov). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In International Conference on Learning Representations.

Dynamic Vision Transformers

Tsinghua University

Not All Images are Worth 16x16 Words!

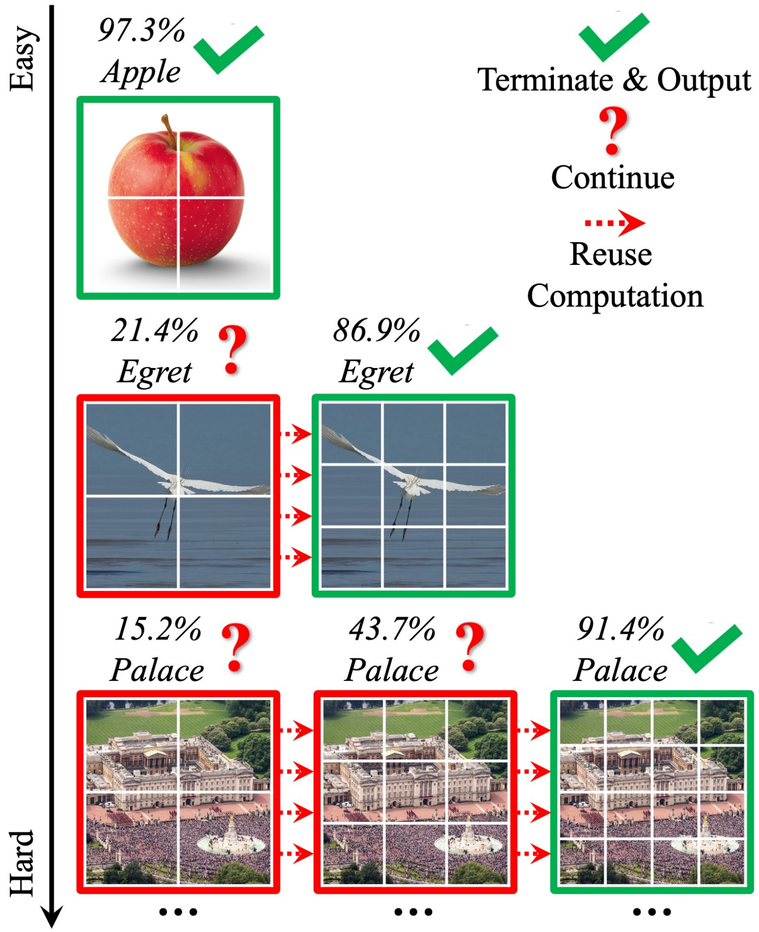
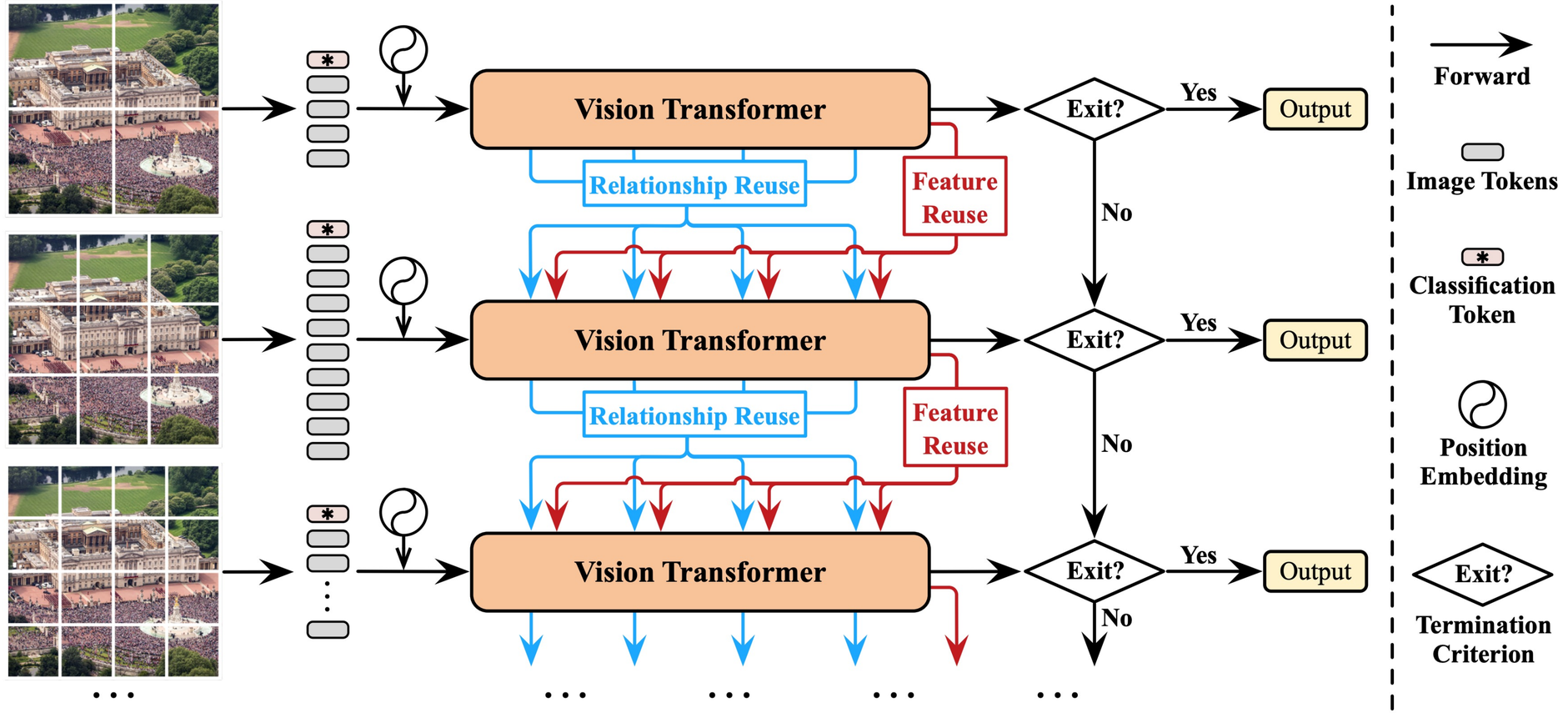


Figure 1: Examples for DVT.

- Wang Y, Huang R, Song S, et al. (2021, September). Not All Images are Worth 16x16 Words: Dynamic Vision Transformers with Adaptive Sequence Length. In Neural Information Processing Systems .

Dynamic Vision Transformers

Dynamic inference with a cascade of ViTs

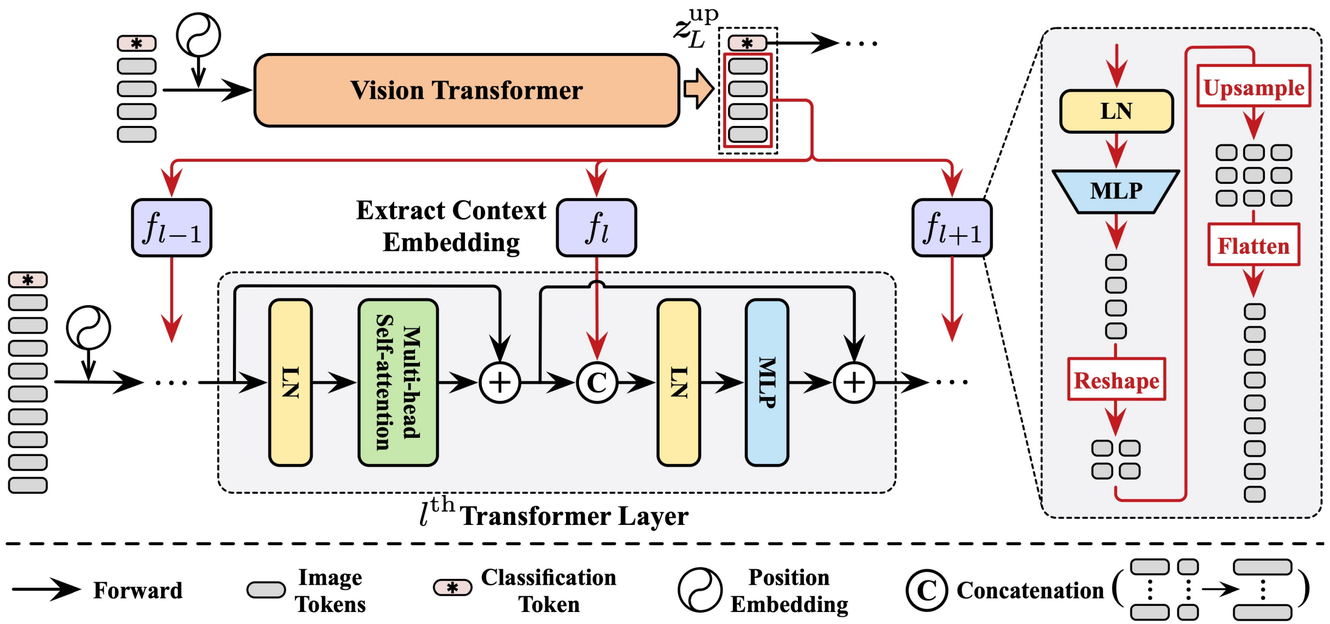


- Wang Y, Huang R, Song S, et al. (2021, September). Not All Images are Worth 16x16 Words: Dynamic Vision Transformers with Adaptive Sequence Length. In Neural Information Processing Systems .

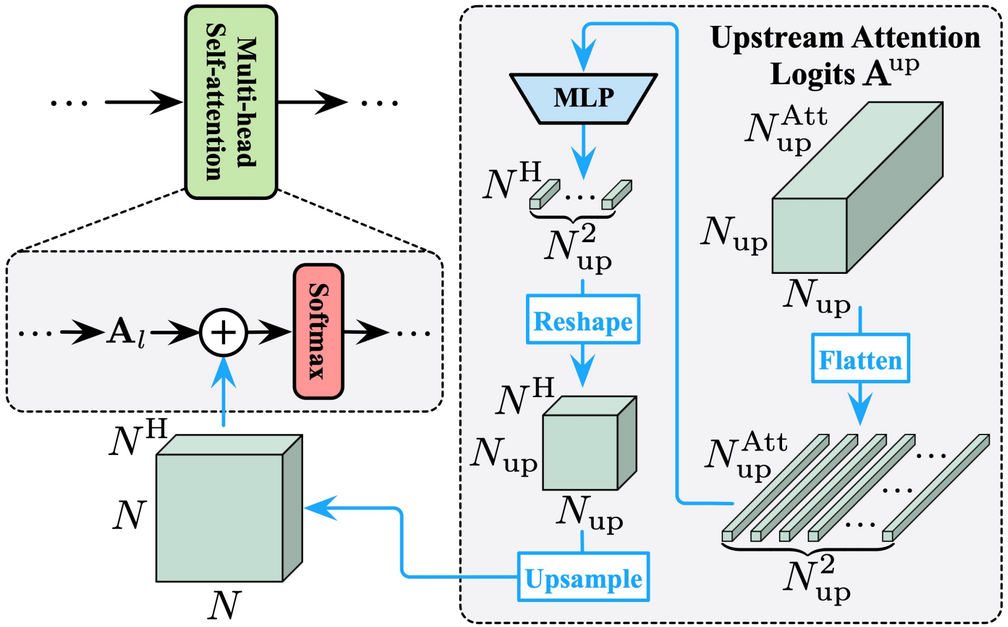
Dynamic Vision Transformers

Tsinghua University

Two key components in dynamic ViTs



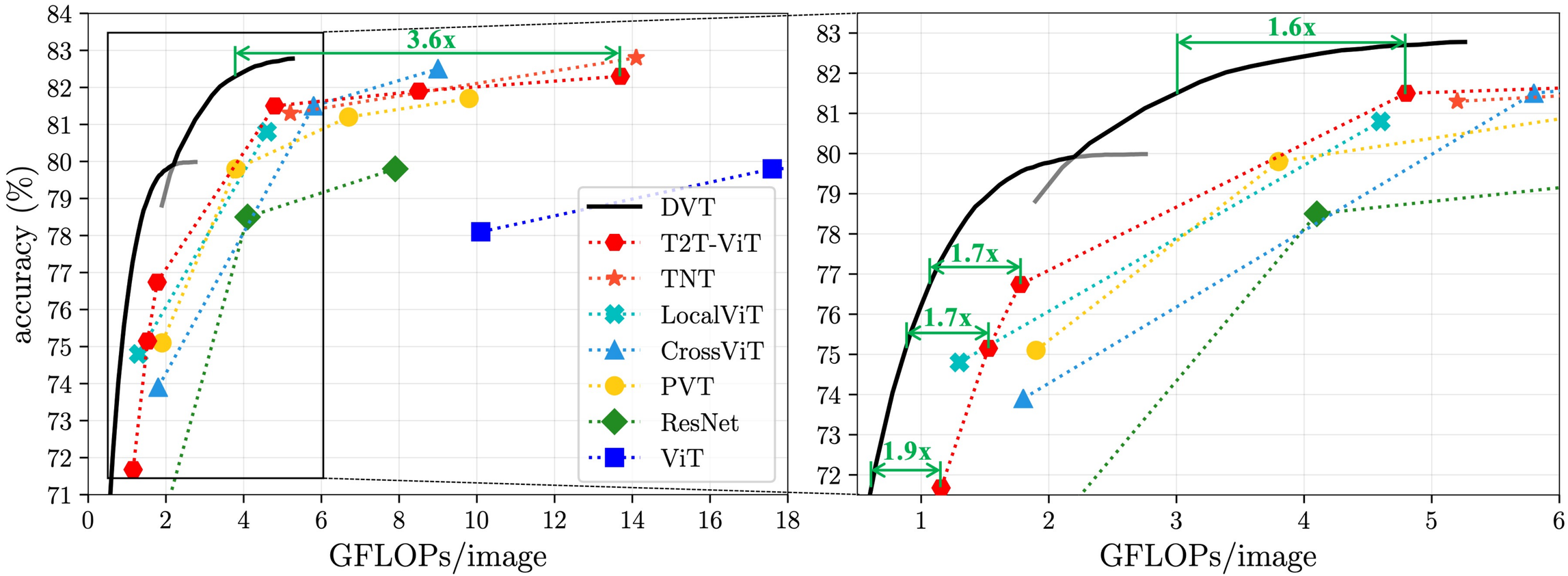
Component (1) Feature Reuse



Component (2) Relationship Reuse

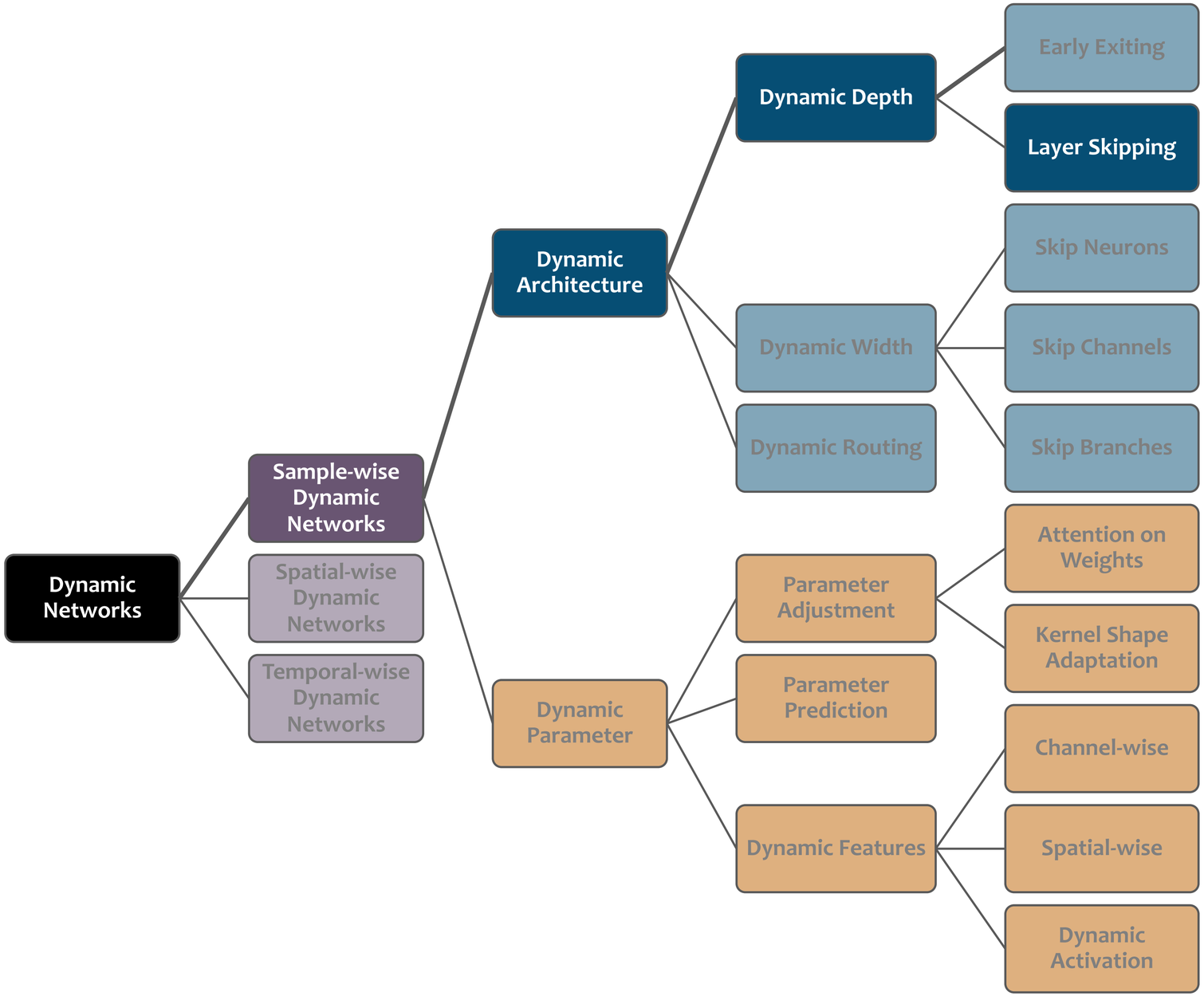
- Wang Y, Huang R, Song S, et al. (2021, September). Not All Images are Worth 16x16 Words: Dynamic Vision Transformers with Adaptive Sequence Length. In Neural Information Processing Systems .

Dynamic Vision Transformers



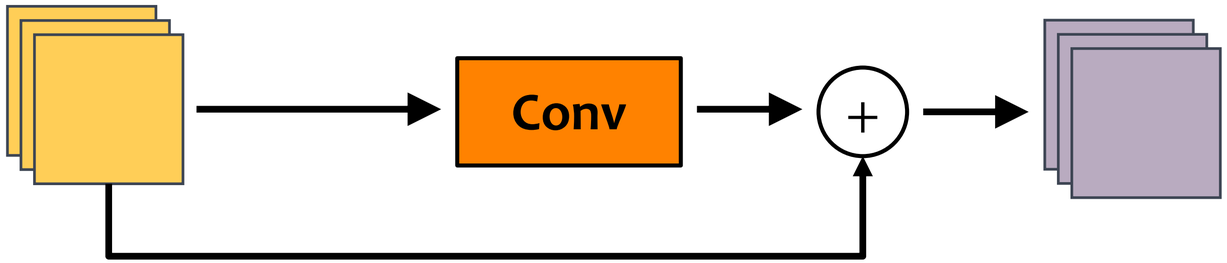
- Wang Y, Huang R, Song S, et al. (2021, September). Not All Images are Worth 16x16 Words: Dynamic Vision Transformers with Adaptive Sequence Length. In Neural Information Processing Systems .

Sample-wise Dynamic Neural Networks

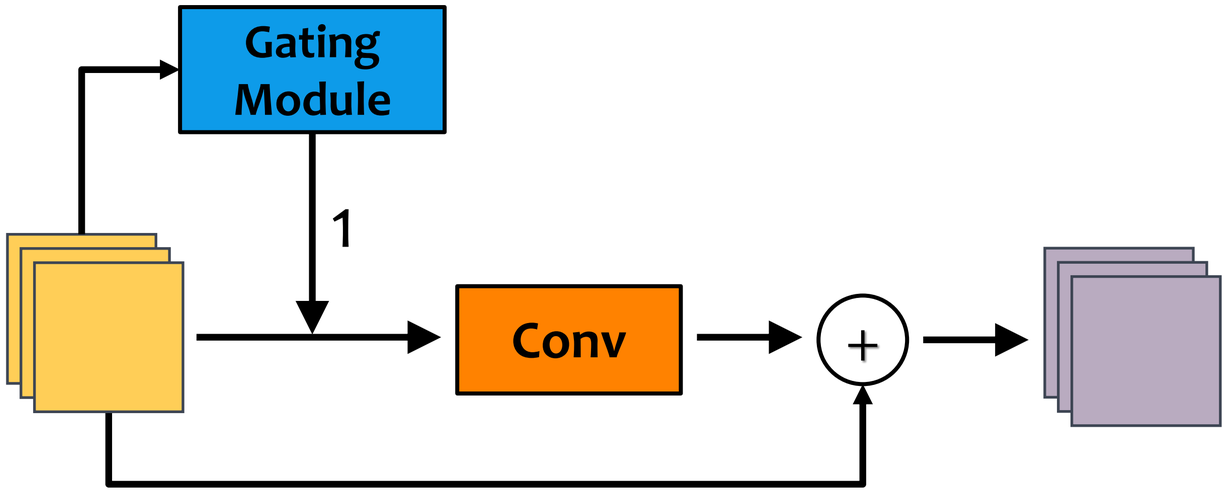


Layer Skipping

A regular residual block

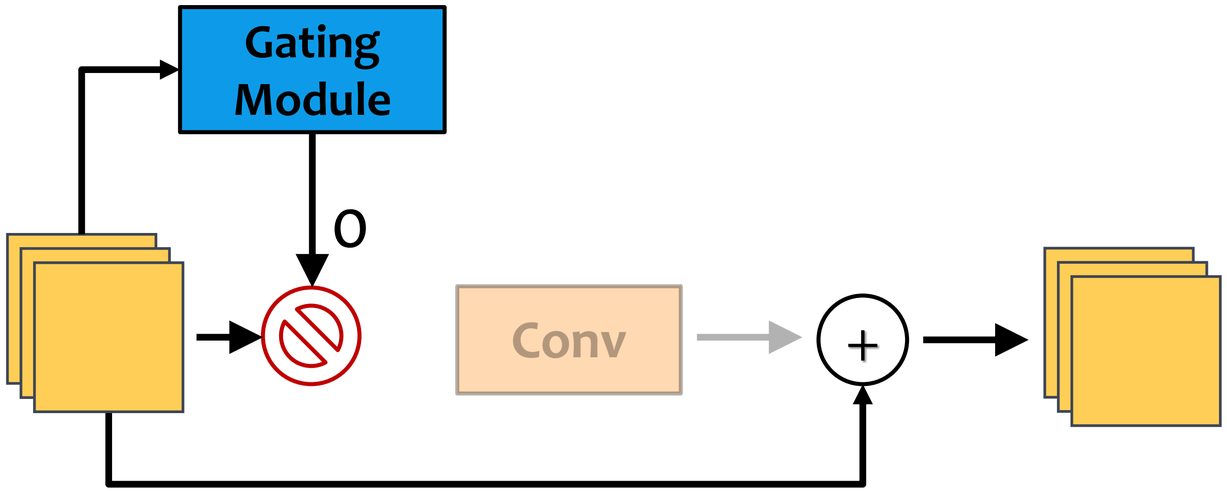


Layer Skipping



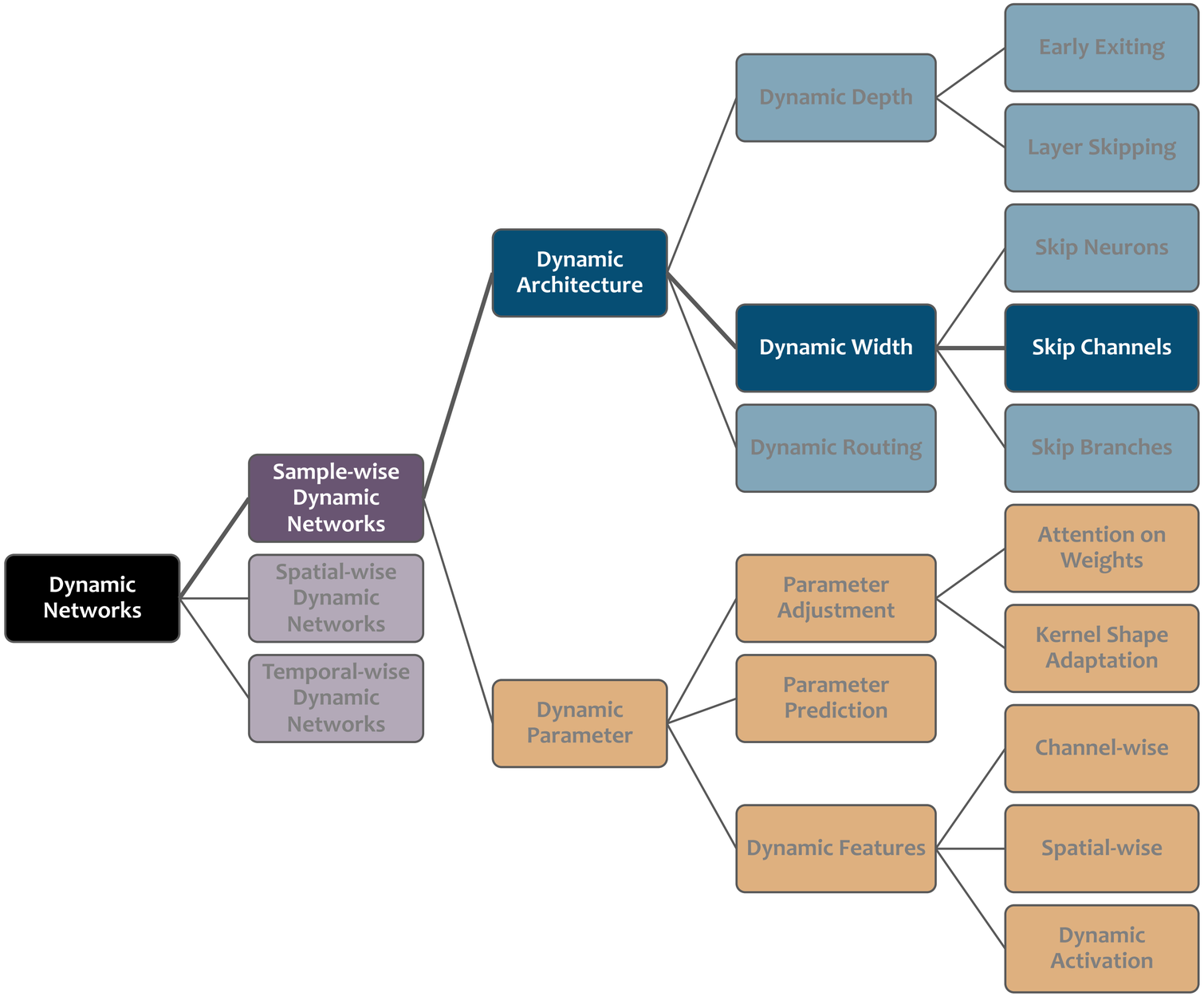
- Wang, X., Yu, F., Dou, Z. Y., Darrell, T., & Gonzalez, J. E. (2018). Skipnet: Learning dynamic routing in convolutional networks. In Proceedings of the European Conference on Computer Vision (ECCV).
- Veit, A., & Belongie, S. (2018). Convolutional networks with adaptive inference graphs. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 3-18).

Layer Skipping

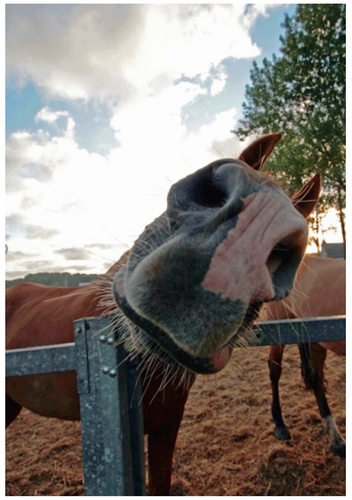


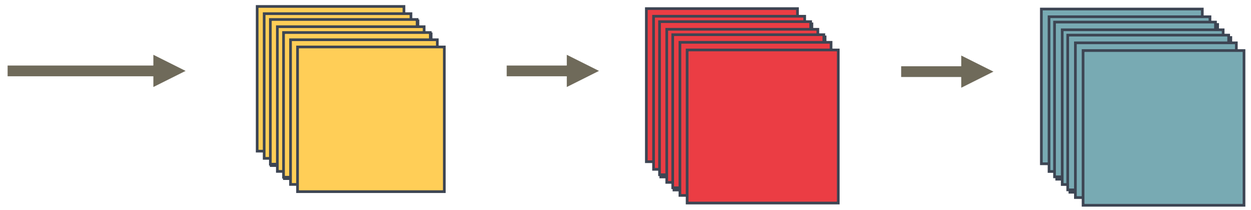
- Wang, X., Yu, F., Dou, Z. Y., Darrell, T., & Gonzalez, J. E. (2018). Skipnet: Learning dynamic routing in convolutional networks. In Proceedings of the European Conference on Computer Vision (ECCV).
- Veit, A., & Belongie, S. (2018). Convolutional networks with adaptive inference graphs. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 3-18).

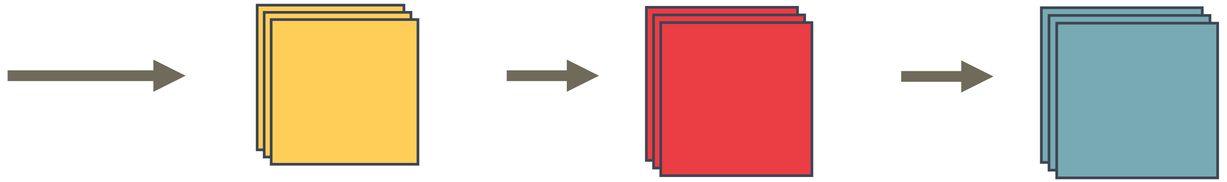
Sample-wise Dynamic Neural Networks



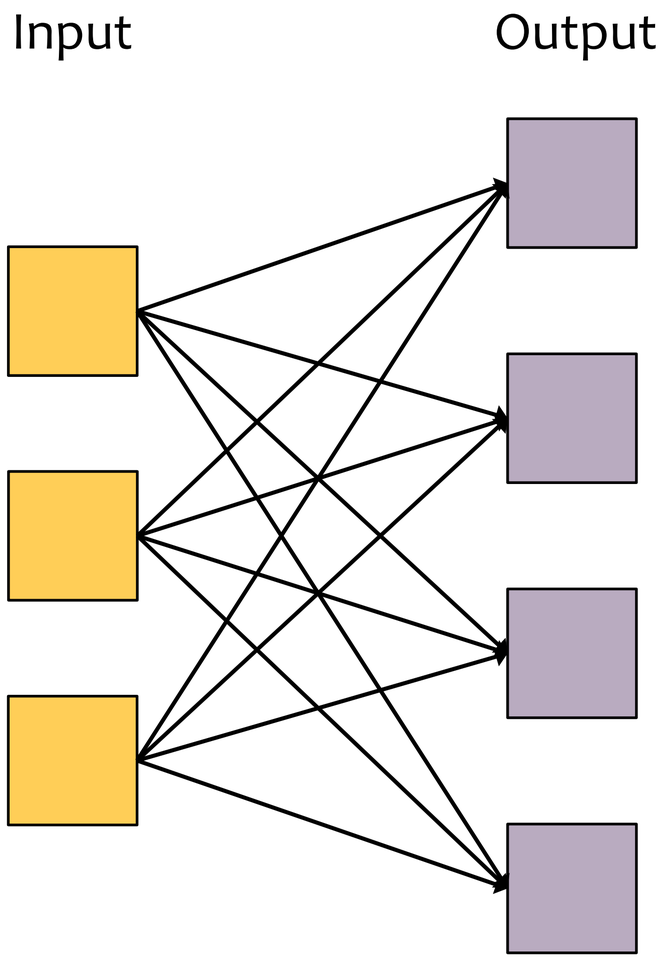
Dynamic Width



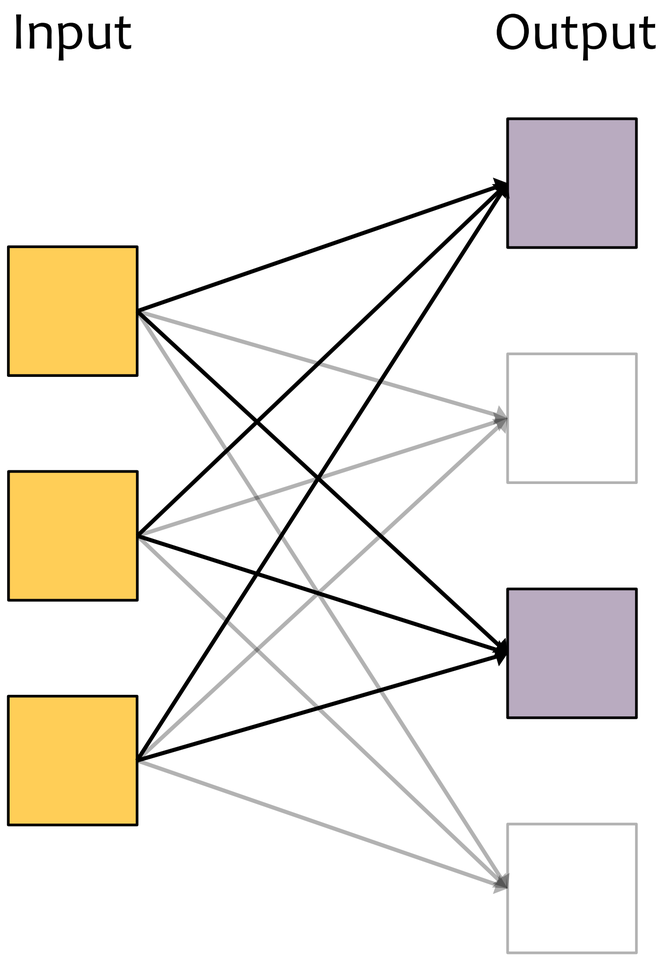




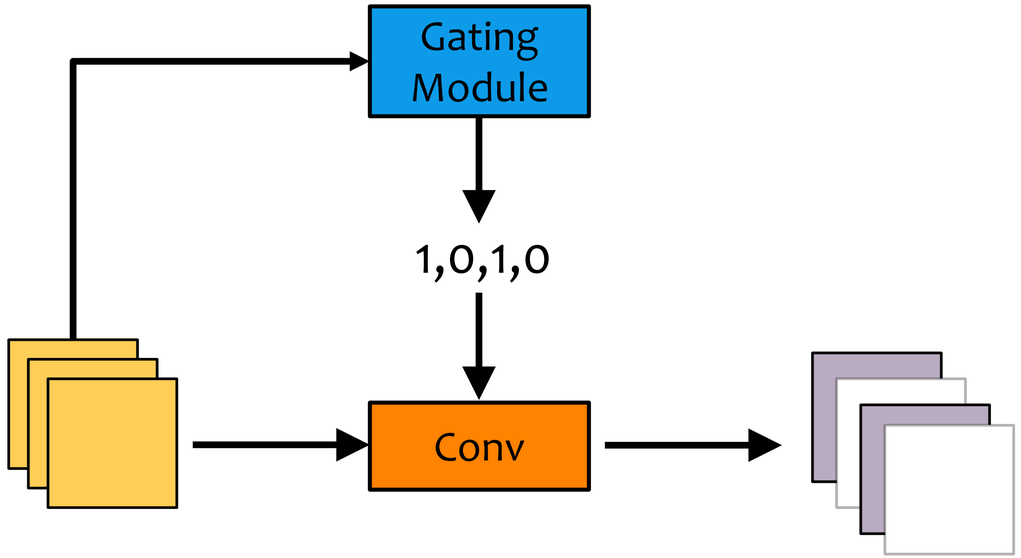
Skip Channels



Skip Channels

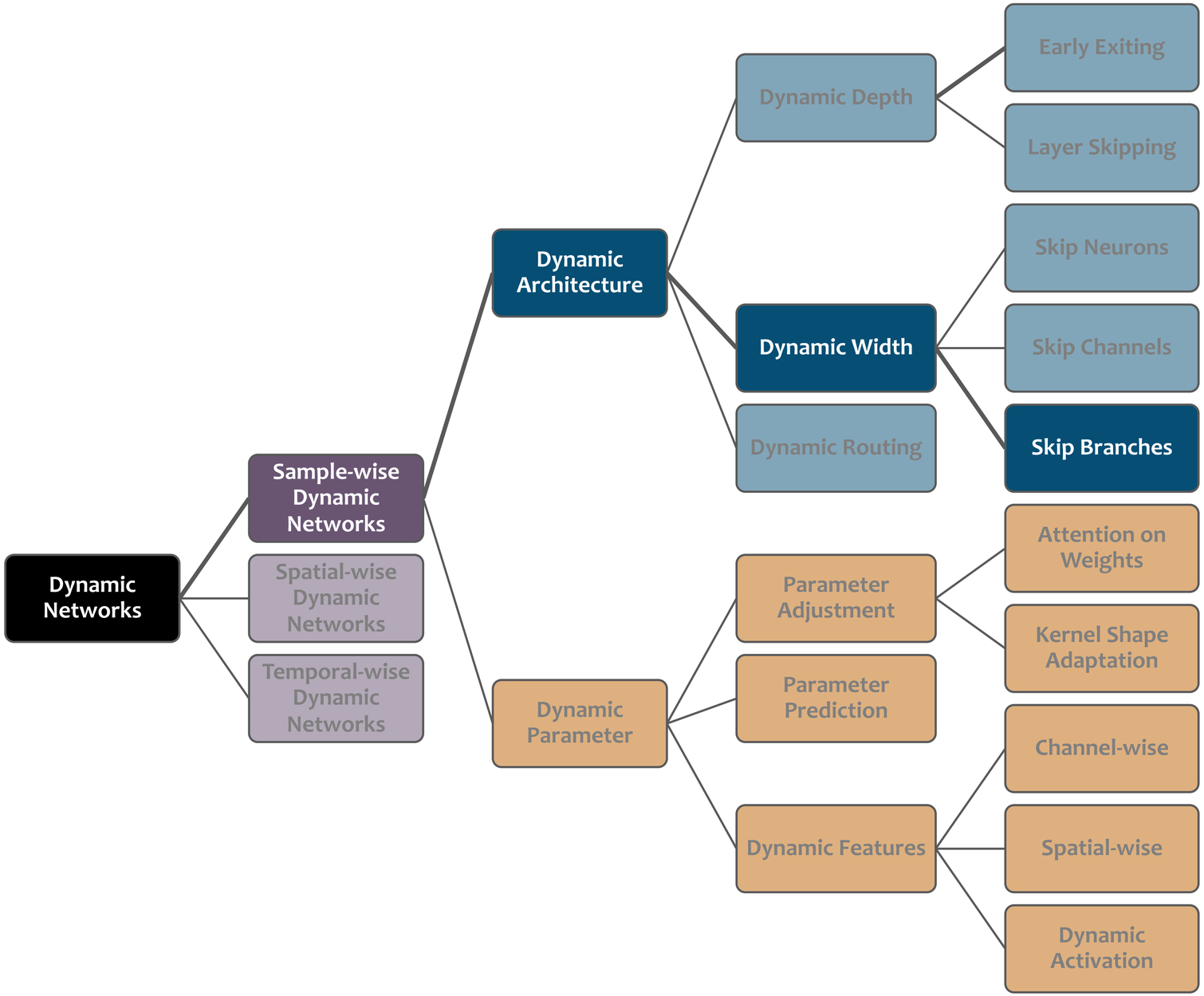


Skip Channels based on Gating Function



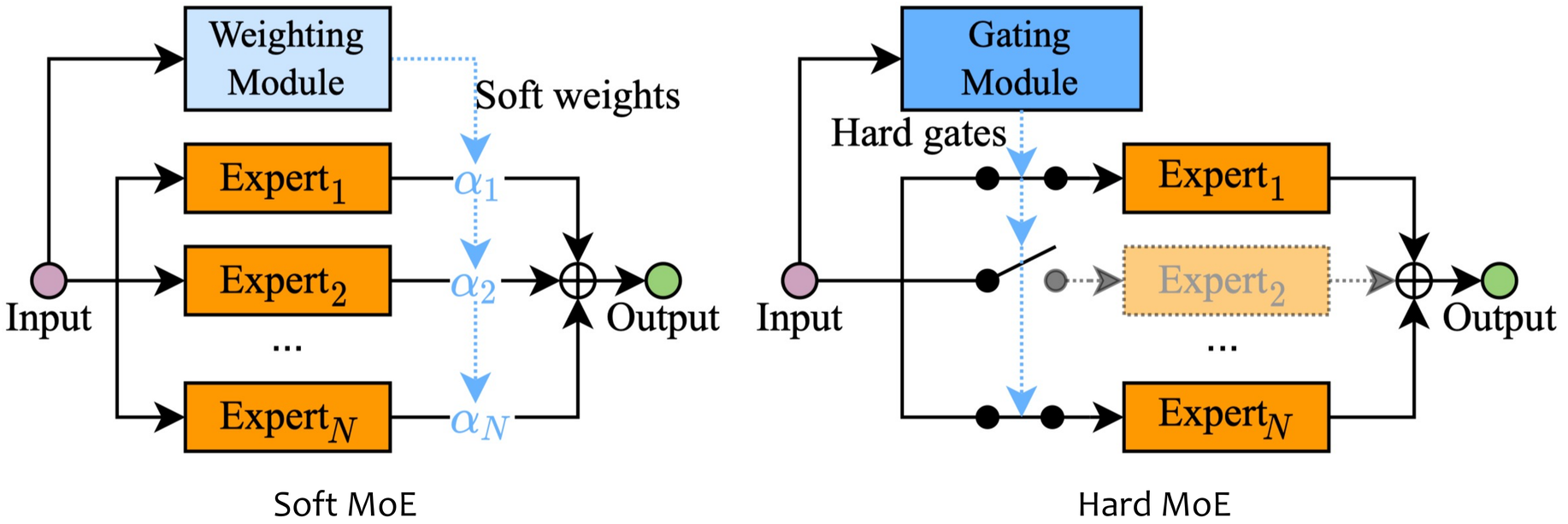
- Lin, J., Rao, Y., Lu, J., & Zhou, J. (2017, December). Runtime neural pruning. In Proceedings of the 31st International Conference on Neural Information Processing Systems.
- Herrmann, C., Bowen, R. S., & Zabih, R. (2020, August). Channel Selection Using Gumbel Softmax. In European Conference on Computer Vision.
- Bejnordi, B. E., Blankevoort, T., & Welling, M. (2019, September). Batch-shaping for learning conditional channel gated networks. In International Conference on Learning Representations.

Sample-wise Dynamic Neural Networks



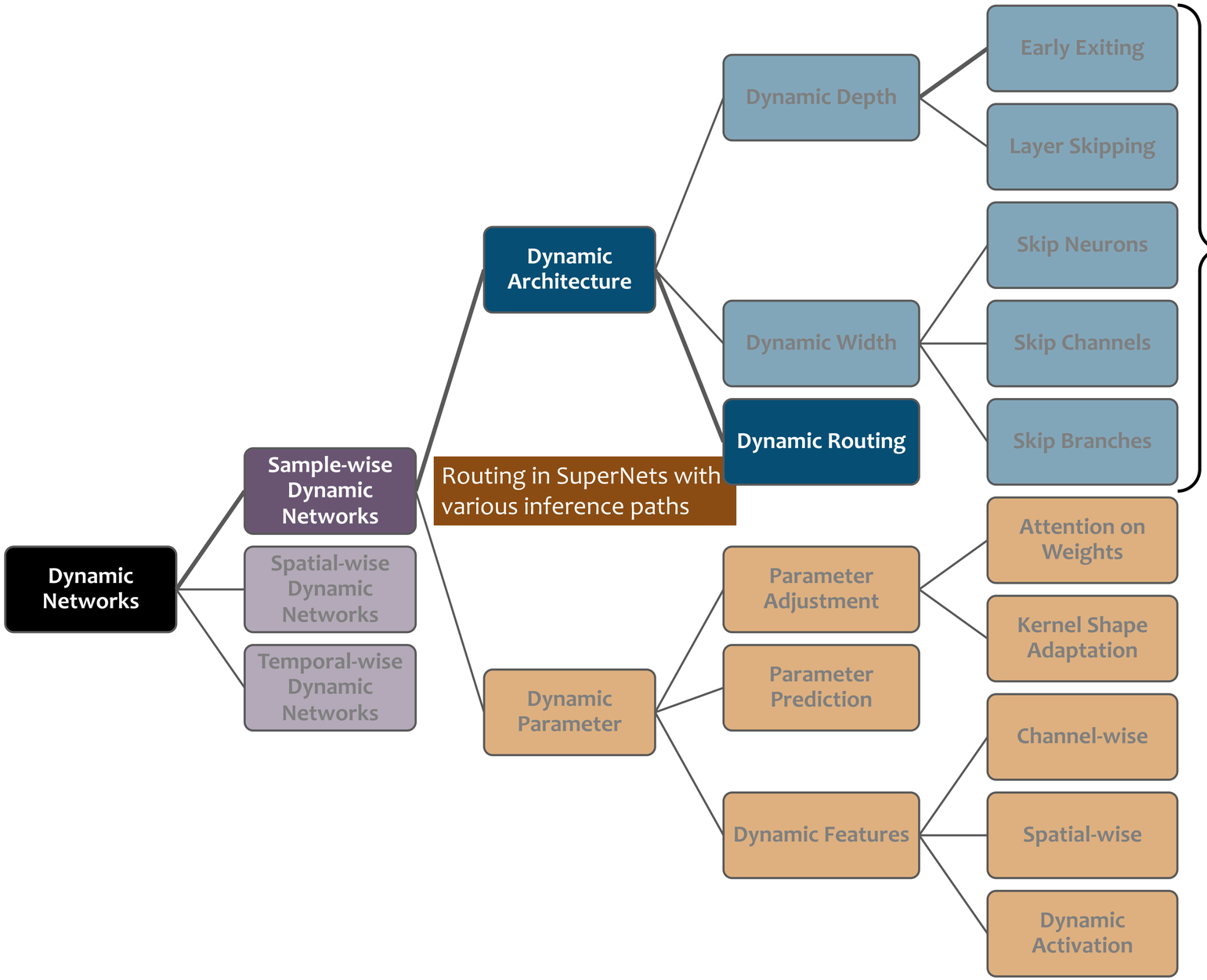
Skip Branches

Mixture of Experts (MoE)

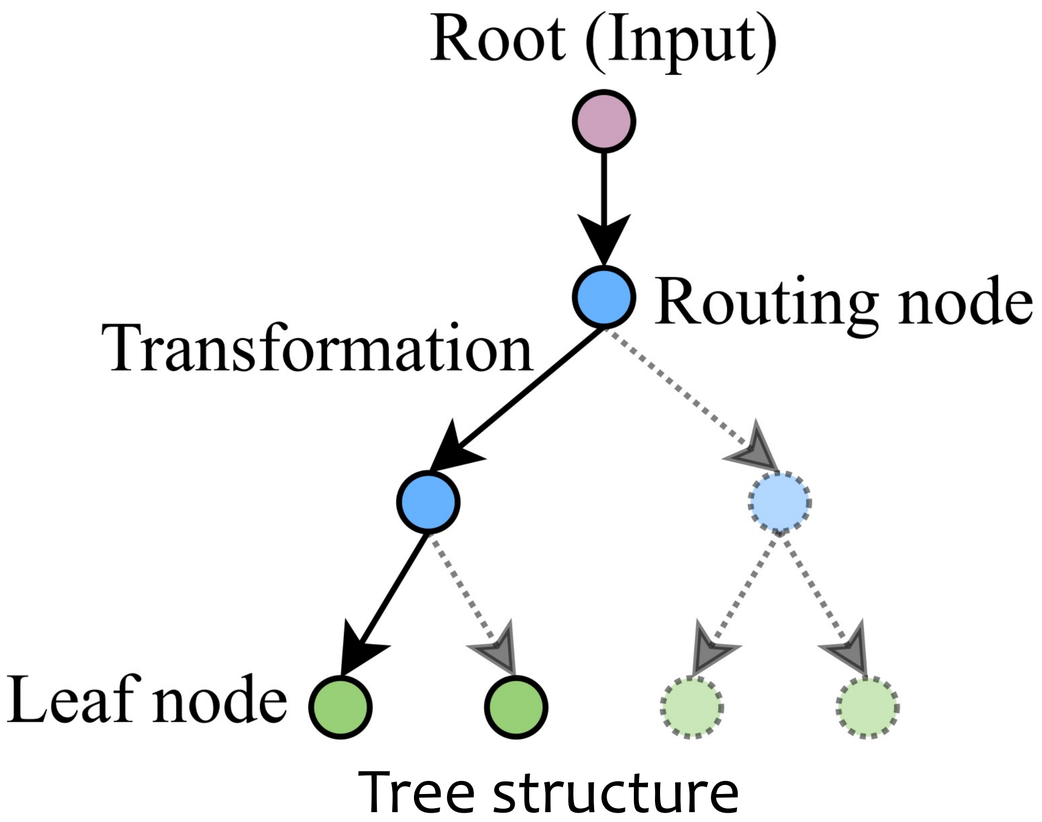


- Mullapudi, R. T., Mark, W. R., Shazeer, N., & Fatahalian, K. (2018). Hydranets: Specialized dynamic architectures for efficient inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 8080-8089).
- Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538.
- Fedus, W., Zoph, B., & Shazeer, N. (2021). Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. arXiv preprint arXiv:2101.03961.

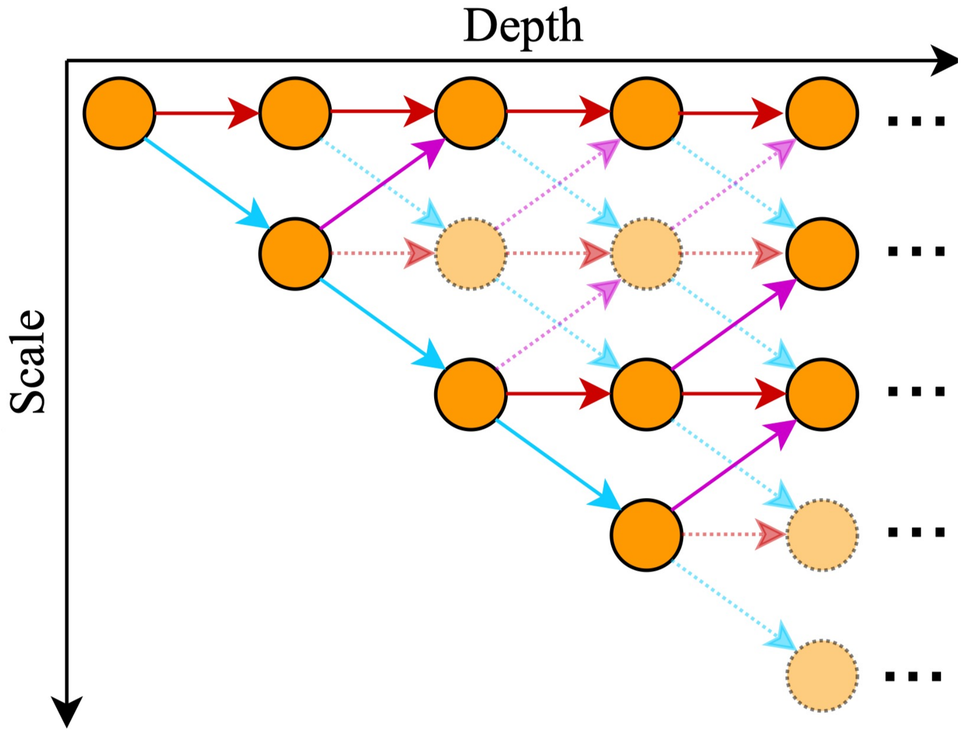
Sample-wise Dynamic Neural Networks



Dynamic Routing in SuperNets



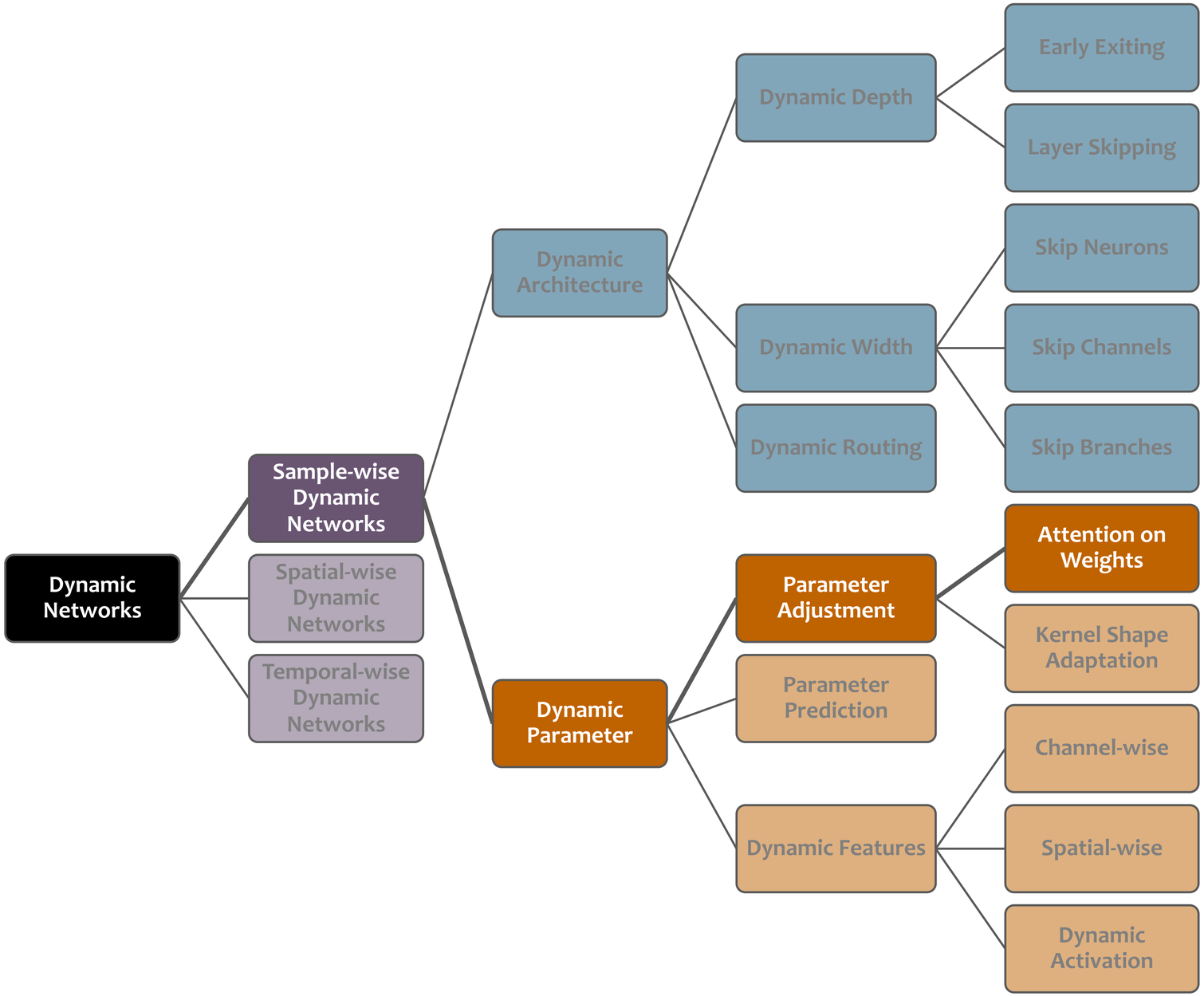
Tree structure



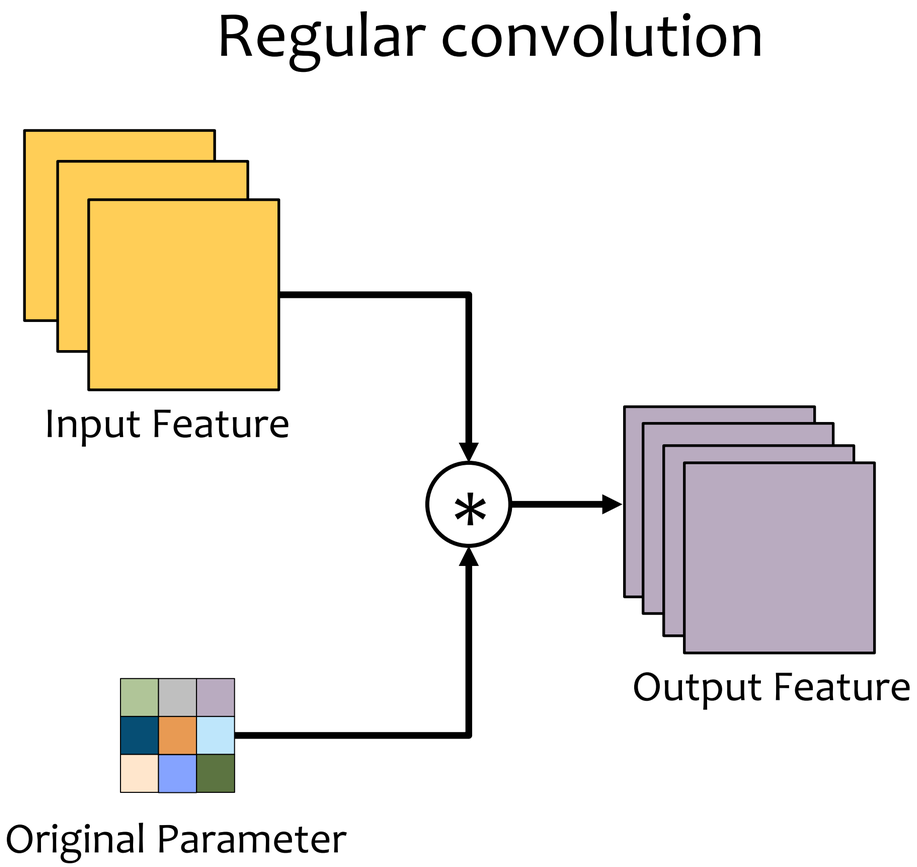
Multi-scale structure

- Tanno, R., Arulkumaran, K., Alexander, D., Criminisi, A., & Nori, A. (2019, May). Adaptive neural trees. In International Conference on Machine Learning (pp. 6166-6175). PMLR.
- Li, Y., Song, L., Chen, Y., Li, Z., Zhang, X., Wang, X., & Sun, J. (2020). Learning dynamic routing for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8553-8562).

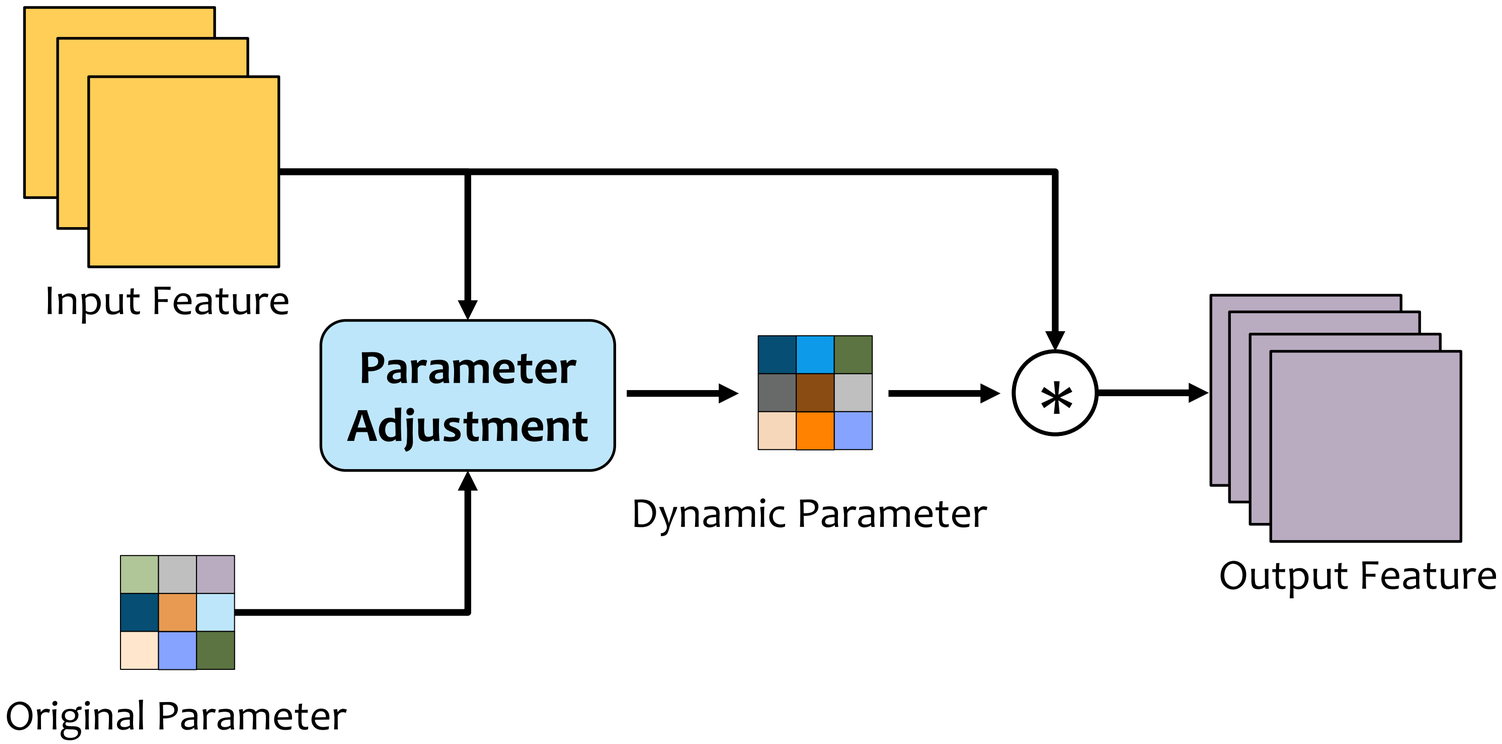
Sample-wise Dynamic Neural Networks



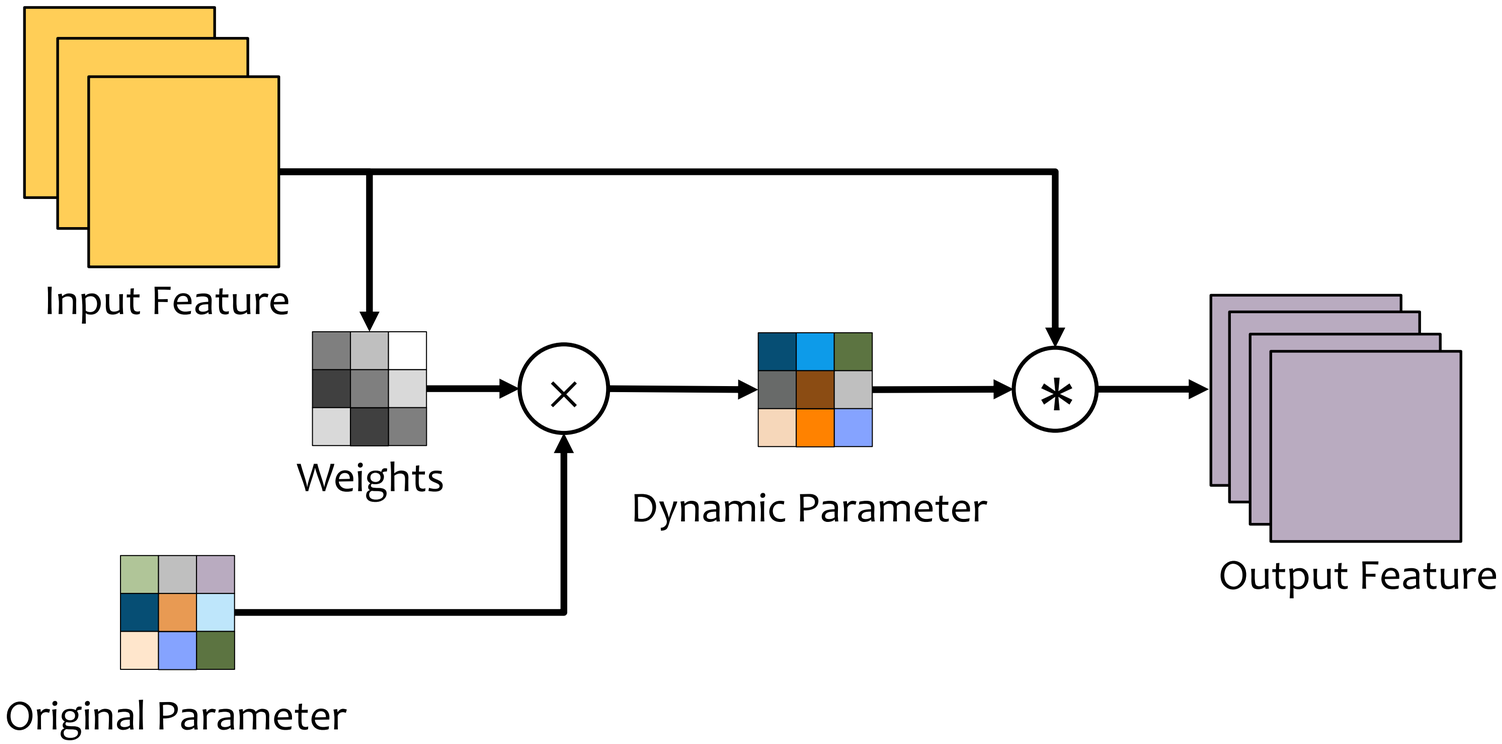
Parameter Adjustment



Parameter Adjustment

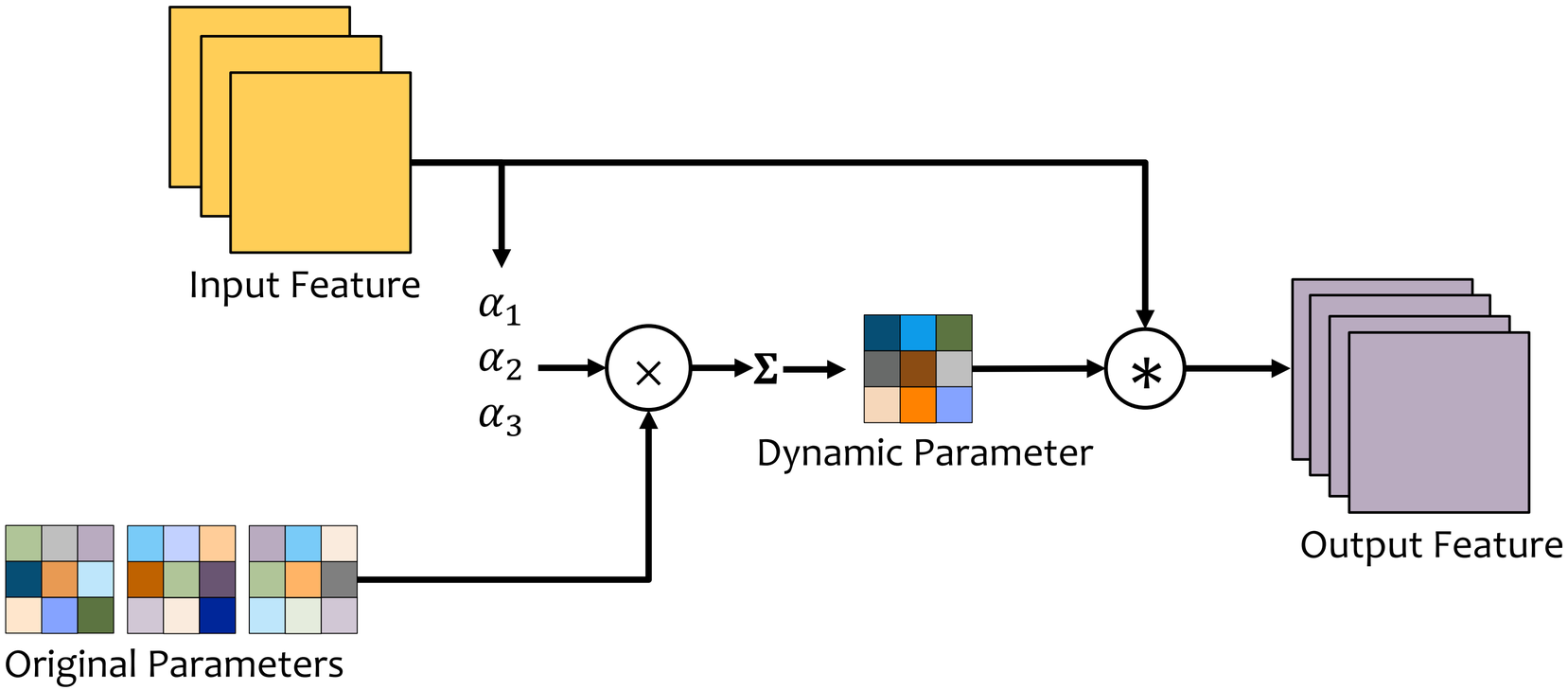


Parameter Adjustment



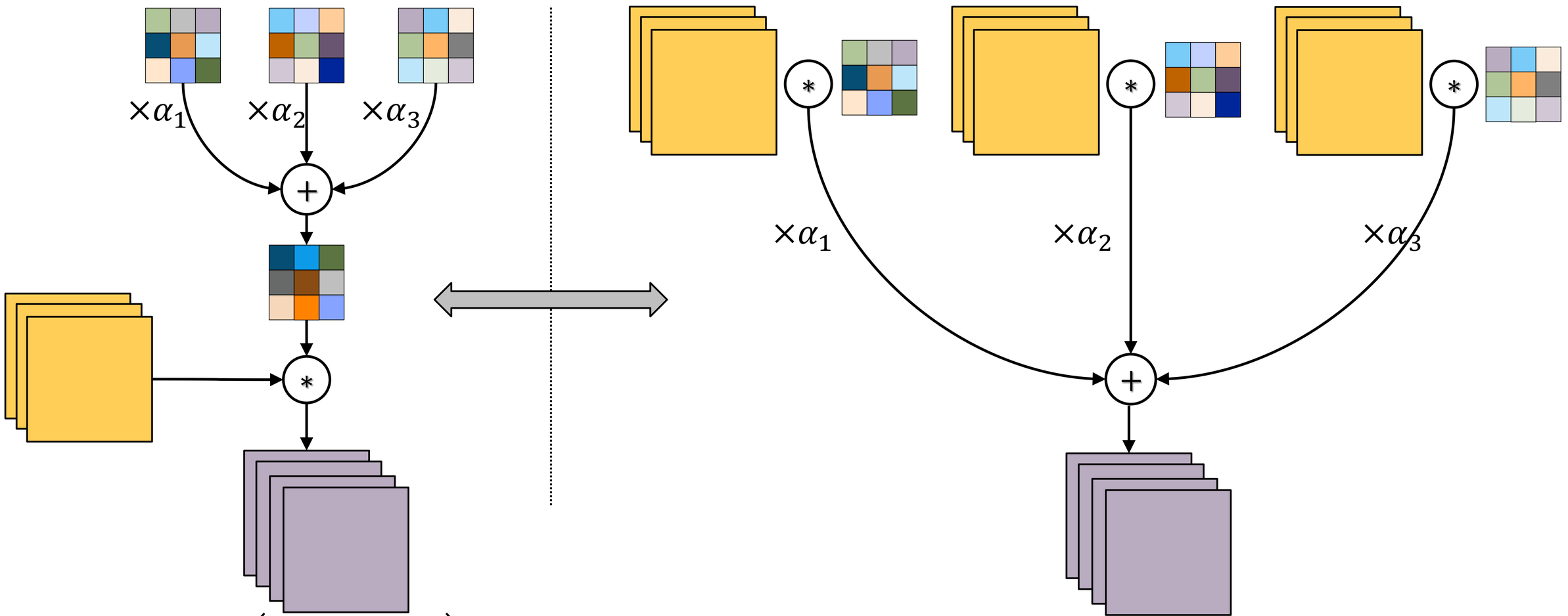
- Su, H., Jampani, V., Sun, D., Gallo, O., Learned-Miller, E., & Kautz, J. (2019). Pixel-adaptive convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11166-11175).

Parameter Adjustment



- Su, H., Jampani, V., Sun, D., Gallo, O., Learned-Miller, E., & Kautz, J. (2019). Pixel-adaptive convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11166-11175).

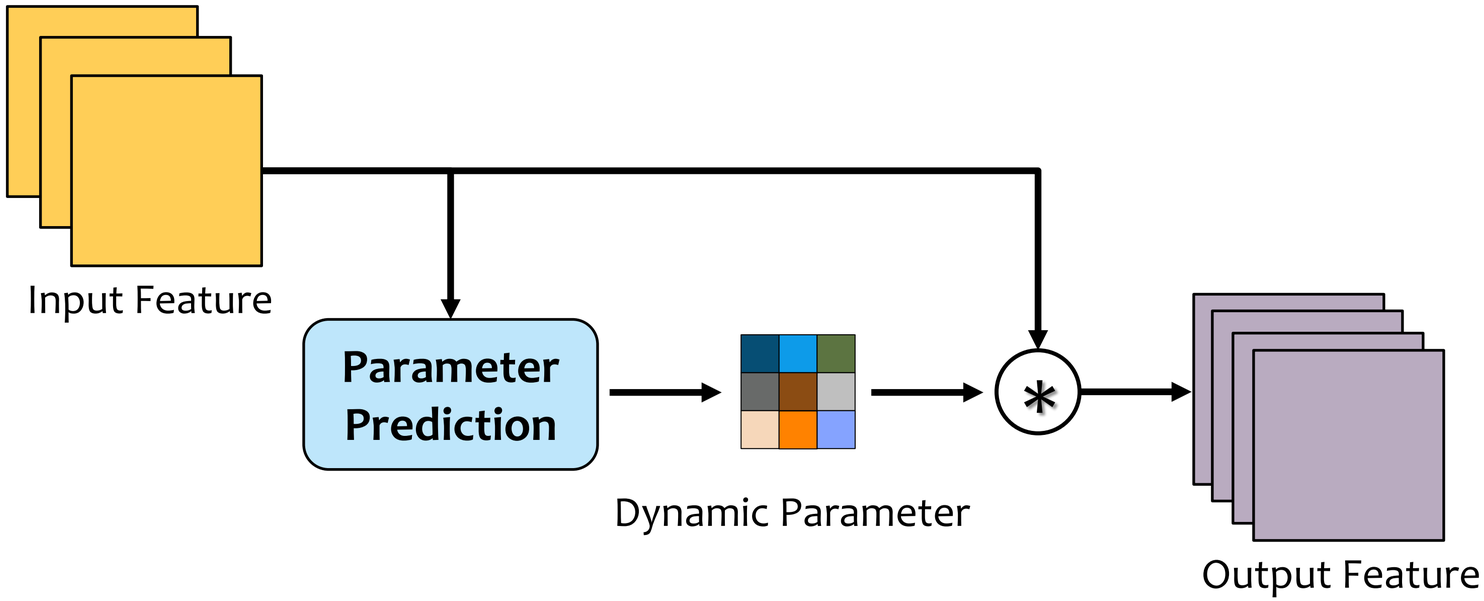
Parameter Adjustment



$$\left(\sum_n \alpha_n \mathbf{W}_n\right) * \mathbf{x} = \sum_n \alpha_n \left(\mathbf{W}_n * \mathbf{x}\right)$$

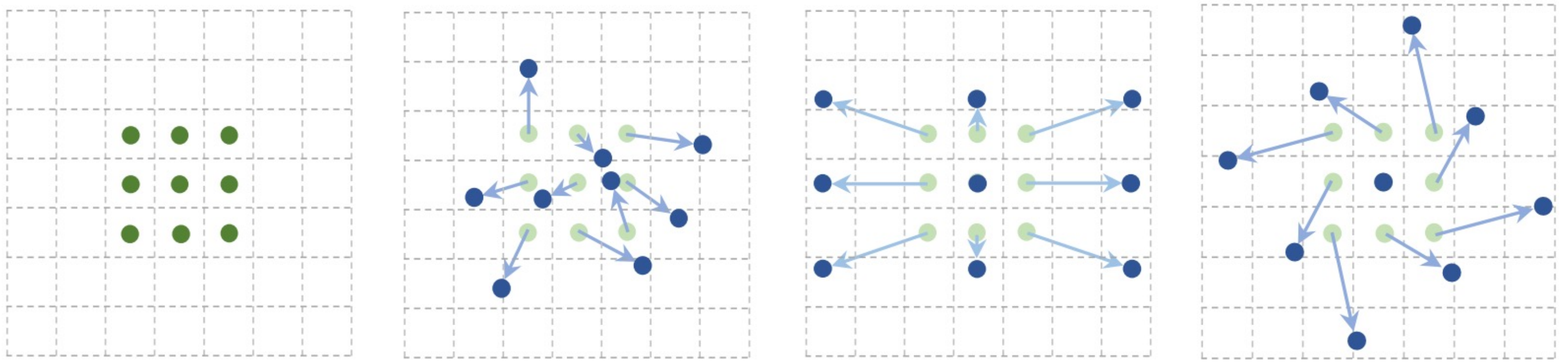
- Yang, B., Bender, G., Le, Q. V., & Ngiam, J. (2019, December). CondConv: conditionally parameterized convolutions for efficient inference. In Proceedings of the 33rd International Conference on Neural Information Processing Systems.

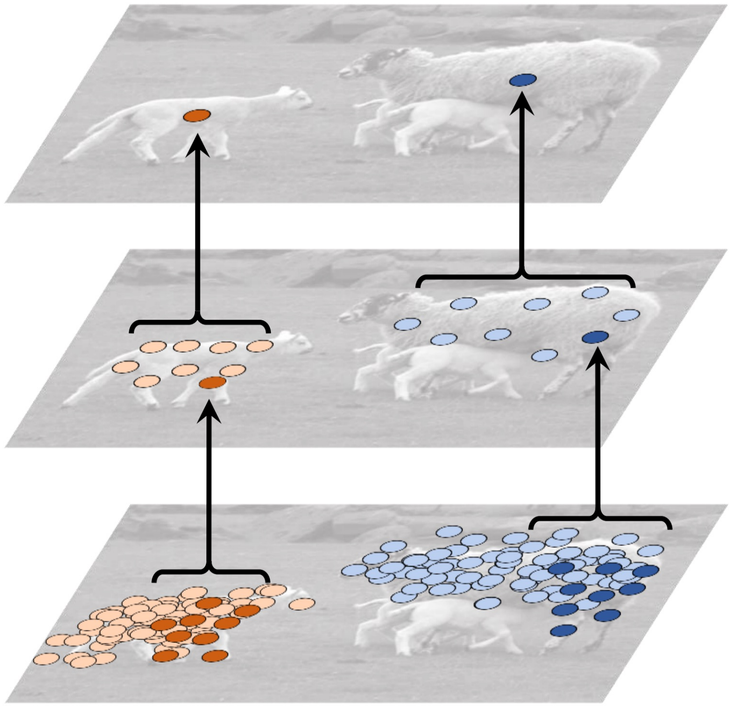
Parameter Prediction



- Ha, David, Andrew M. Dai, and Quoc V. Le. "HyperNetworks." (2016).
- Jia, Xu, et al. "Dynamic filter networks." NeurIPS (2016): 667-675.
- Ma, Ningning, et al. "Weightnet: Revisiting the design space of weight networks." in ECCV (2020).

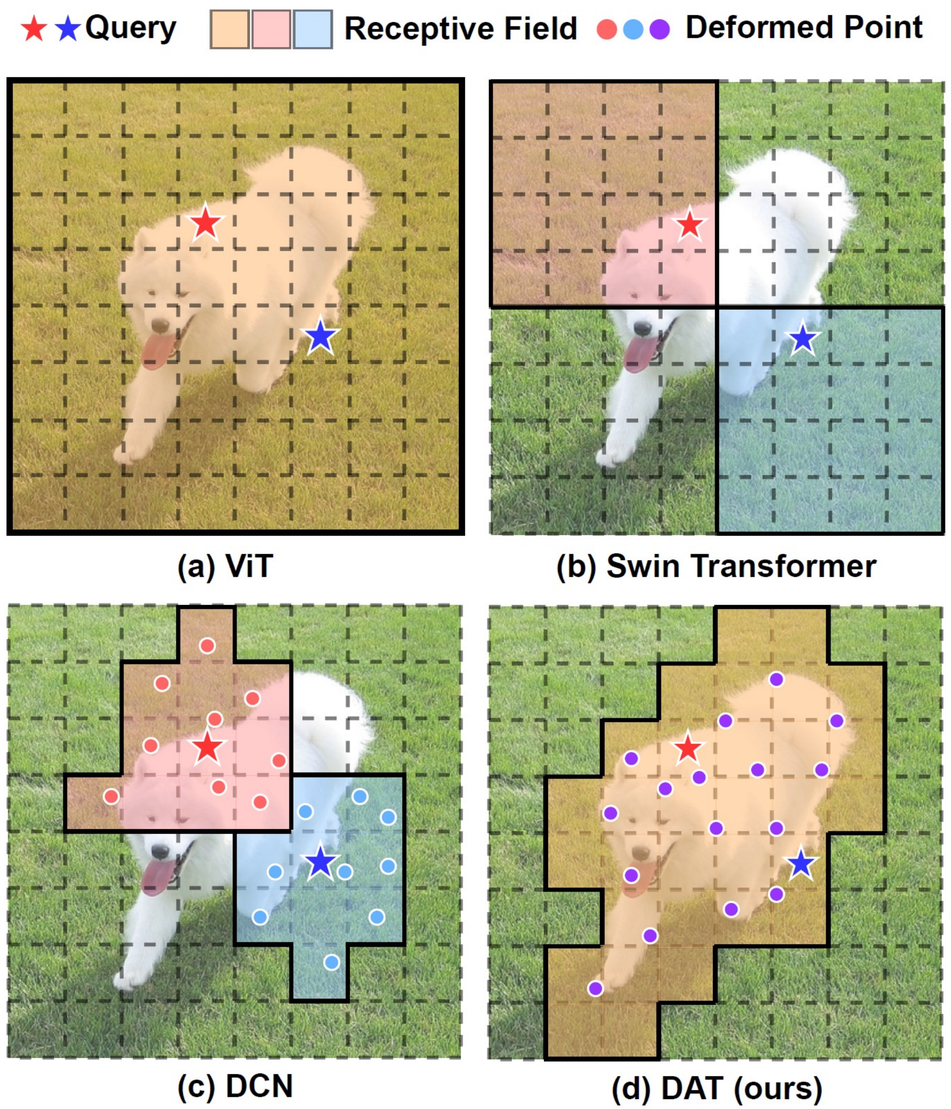
Kernel Shape Adaptation





- Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). Deformable convolutional networks. In Proceedings of the IEEE international conference on computer vision.
- Zhu, X., Hu, H., Lin, S., & Dai, J. (2019). Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

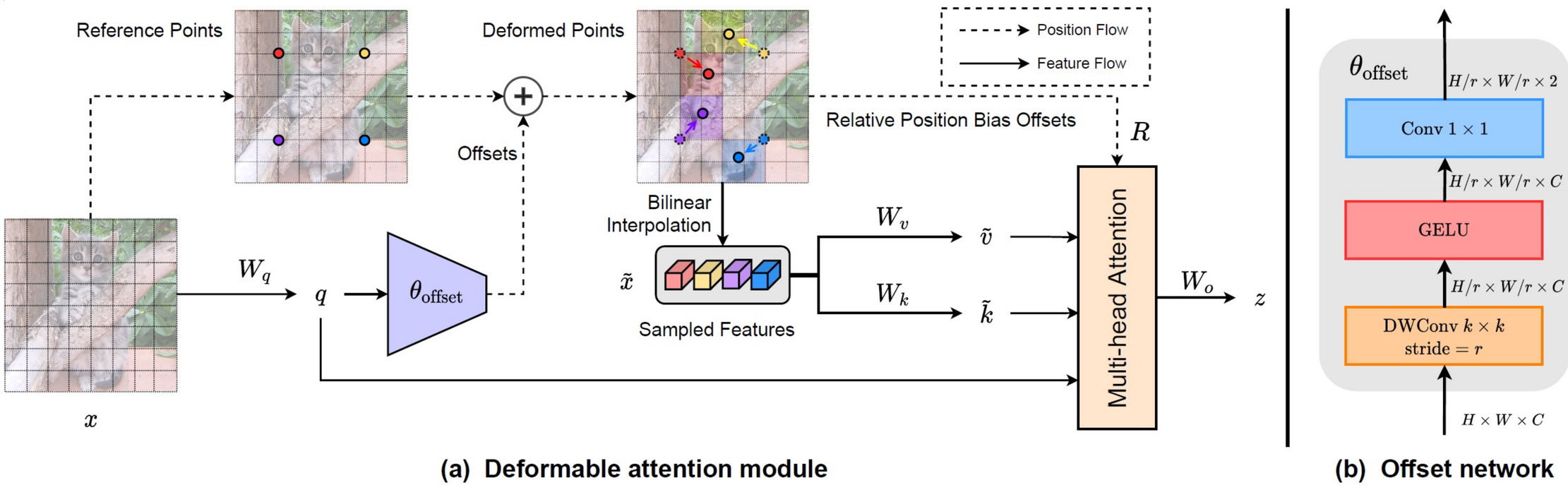
Vision Transformer with Deformable Attention



- Hand crafted attention pattern V.S. data-dependent pattern.
- Deformable attention with shared keys to all queries.
- More flexibility with acceptable memory consumption.

- Xia, Z., Pan, X., Song, S., Li L., Huang, G. Vision Transformer with Deformable Attention. Axiv Preprint, 2022

Vision Transformer with Deformable Attention

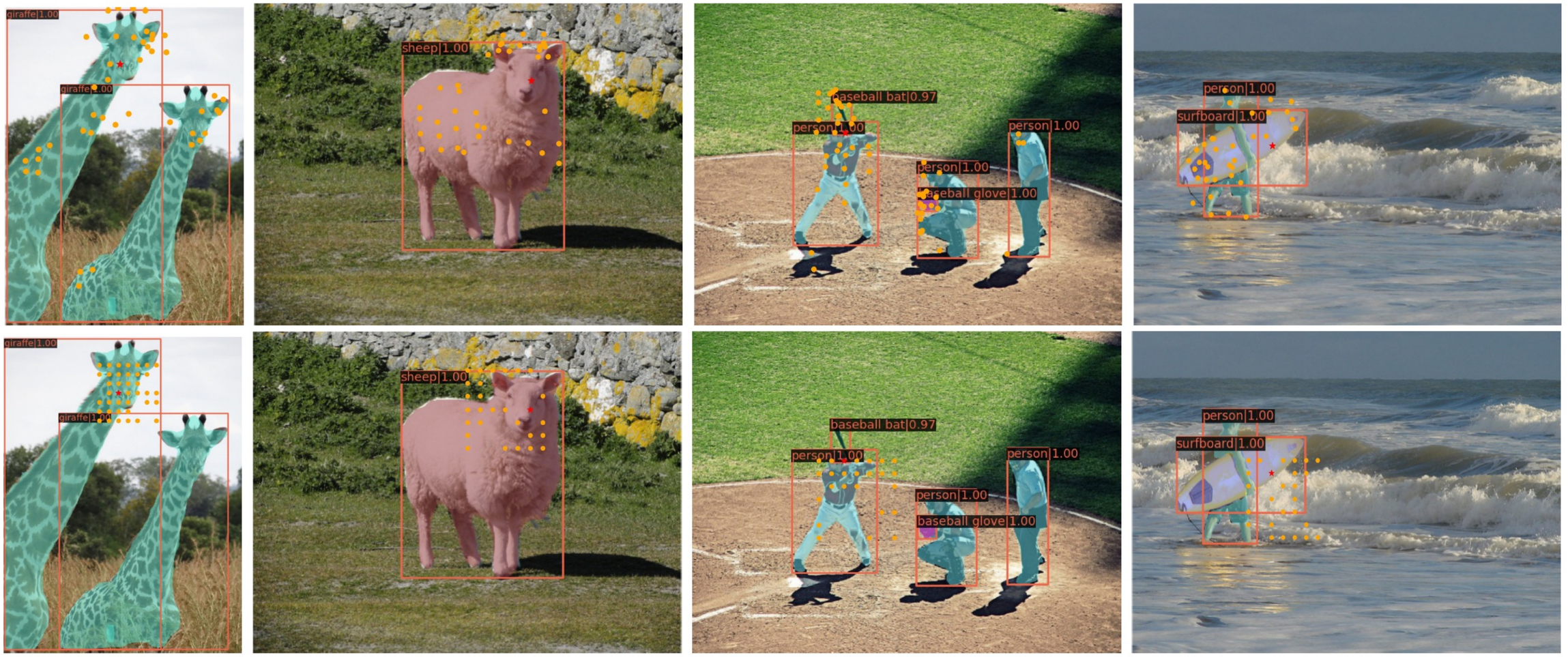


(a) Deformable attention module (b) Offset network

1. Compute offsets from queries.
2. Sample features according to deformed locations.
3. Compute deformed attention.

- Xia, Z., Pan, X., Song, S., Li L., Huang, G. Vision Transformer with Deformable Attention. Axiv Preprint, 2022

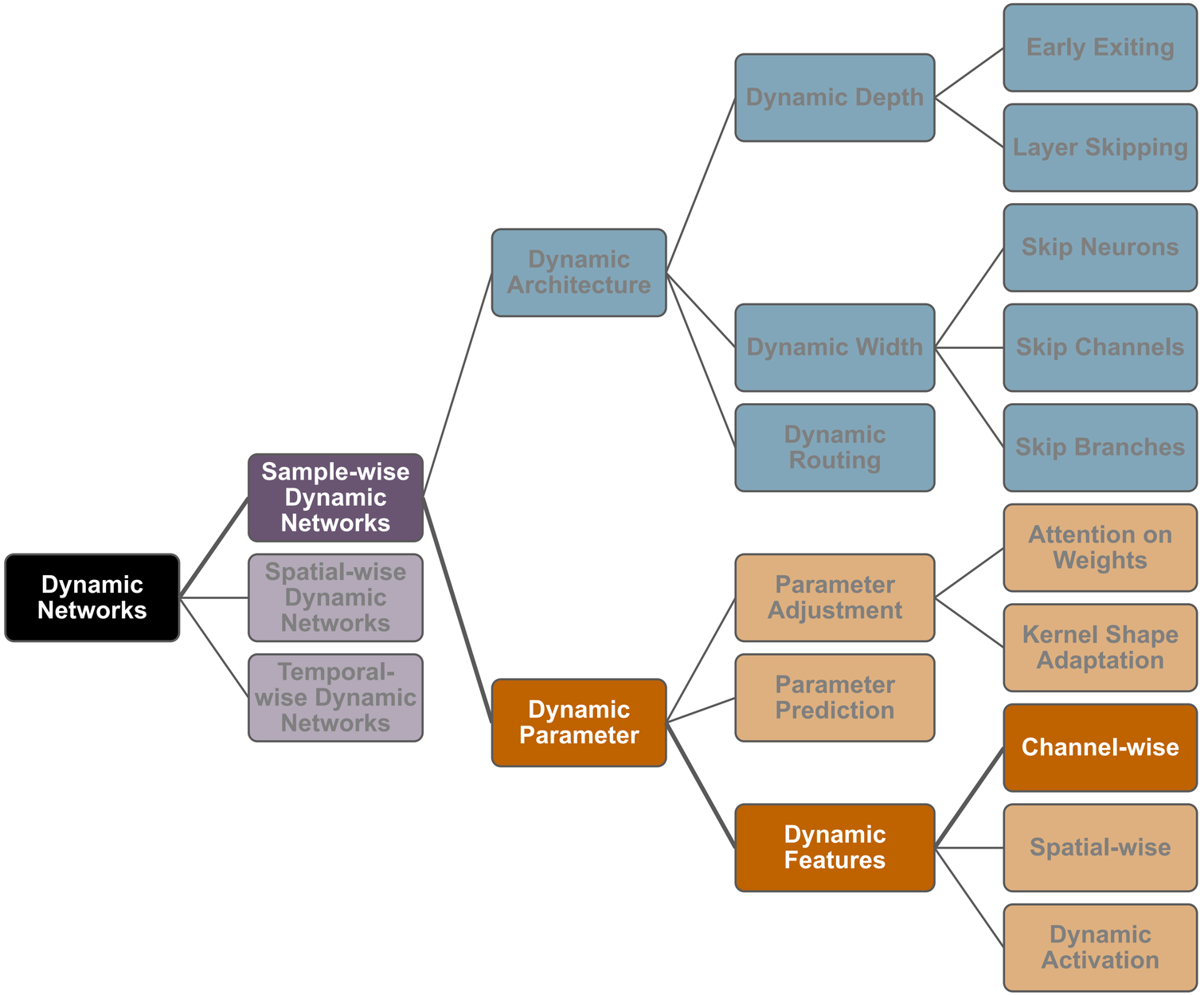
Vision Transformer with Deformable Attention



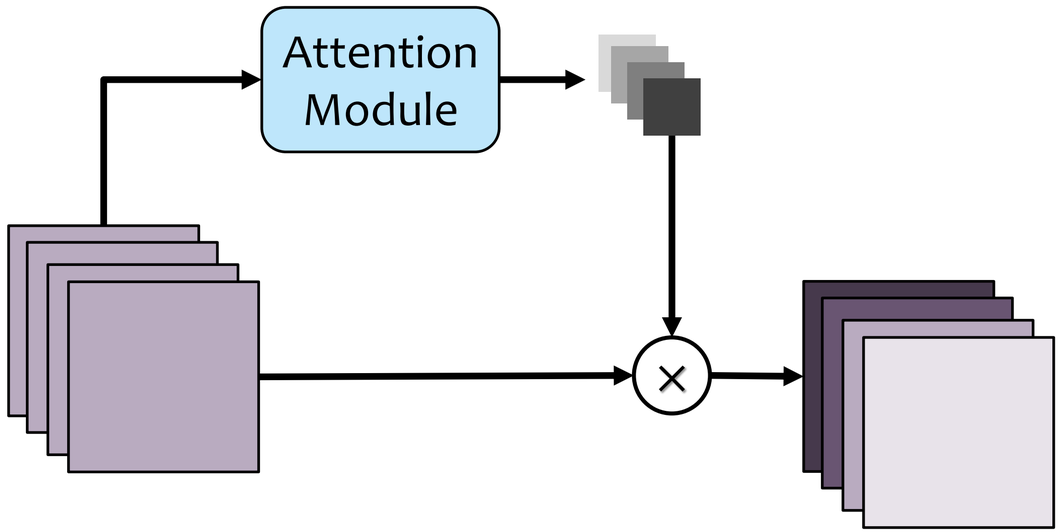
Visualizations of higher attention keys w.r.t. given query (top: ours, bottom: Swin Transformer).

- Xia, Z., Pan, X., Song, S., Li L., Huang, G. Vision Transformer with Deformable Attention. Axiv Preprint, 2022

Sample-wise Dynamic Neural Networks



Channel-wise Attention



Original Output Feature

Dynamic Feature

$$(\mathbf{x} * \mathbf{W}) \otimes \boldsymbol{\alpha} = \mathbf{x} * (\mathbf{W} \otimes \boldsymbol{\alpha})$$

Dynamic Features

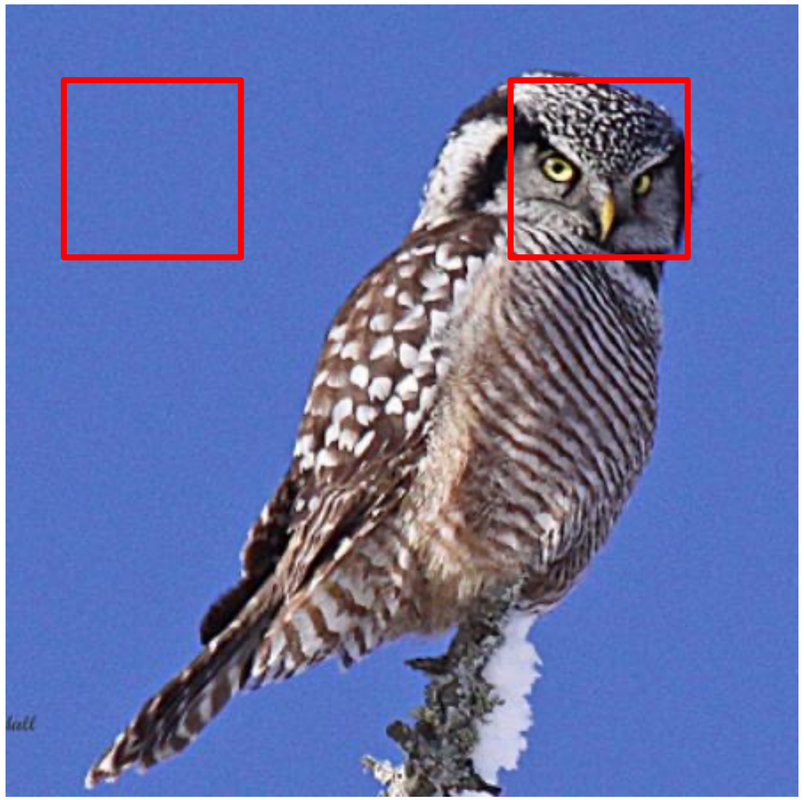
Dynamic Weights

- Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

*From **Sample Adaptive** to **Spatial Adaptive***

Most conventional networks perform the same computation across different ***spatial locations of an image.***

Less Informative

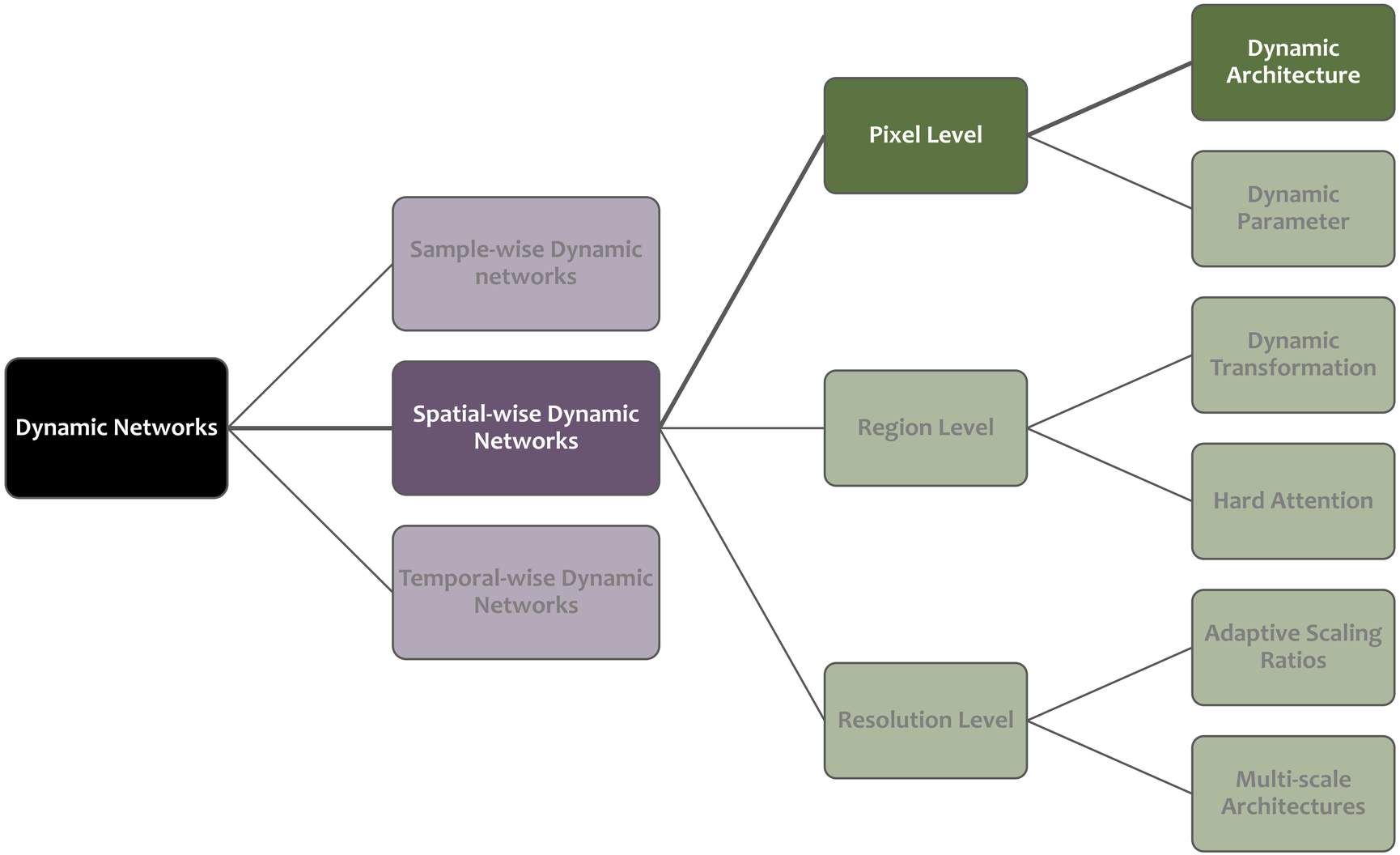


More Informative

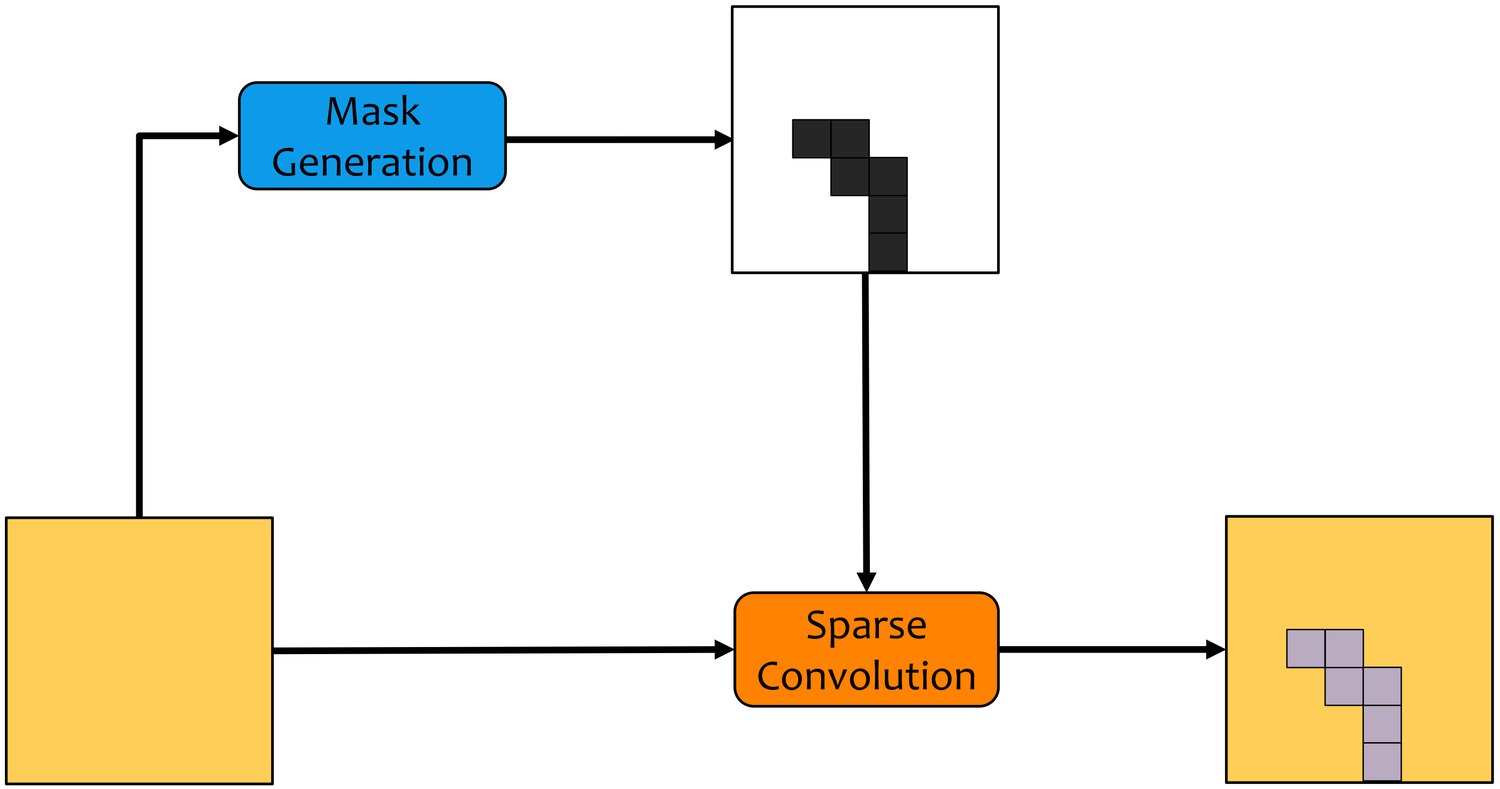
Overview

- Introduction
- Sample-wise Dynamic Networks
- **Spatial-wise Dynamic Networks**
- Temporal-wise Dynamic Networks
- Inference & Training
- Applications
- Discussion

Spatial-wise Dynamic Neural Networks



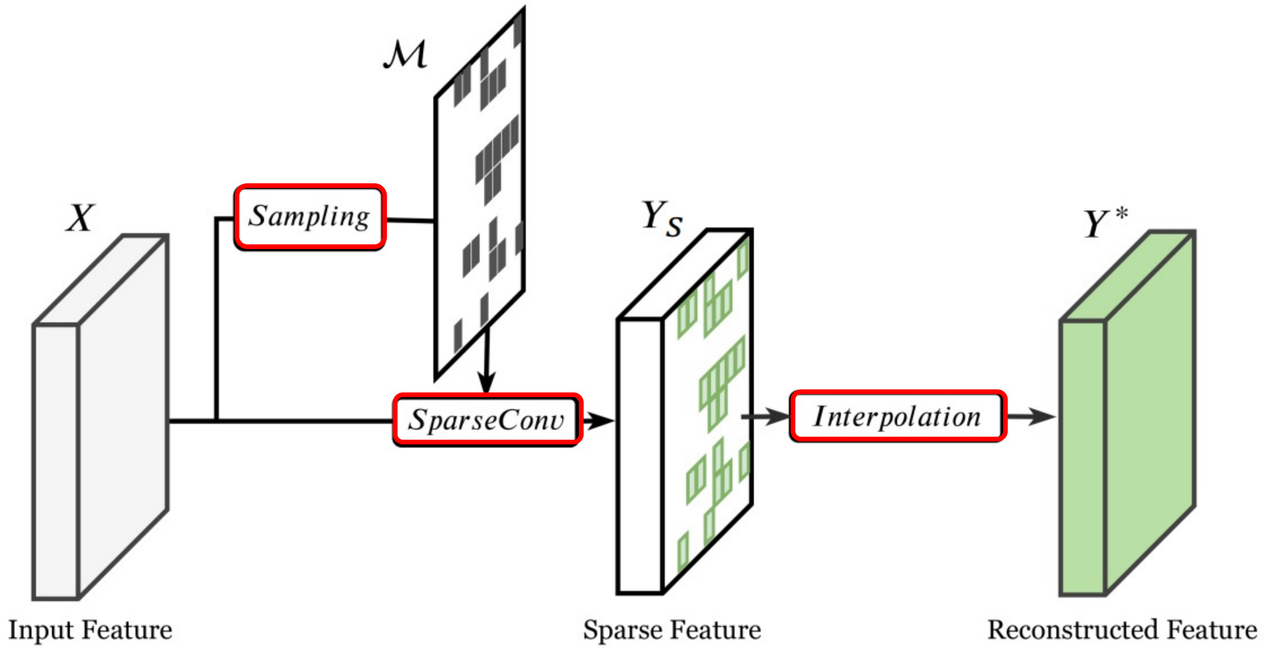
Pixel-level Dynamic Network



- Dong, X., Huang, J., Yang, Y., & Yan, S. (2017). More is less: A more complicated network with less inference complexity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5840-5848).
- Verelst, T., & Tuytelaars, T. (2020). Dynamic convolutions: Exploiting spatial sparsity for faster inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2320-2329).
- Xie, Z., Zhang, Z., Zhu, X., Huang, G., & Lin, S. (2020, August). Spatially adaptive inference with stochastic feature sampling and interpolation. In European Conference on Computer Vision (pp. 531-548). Springer, Cham.

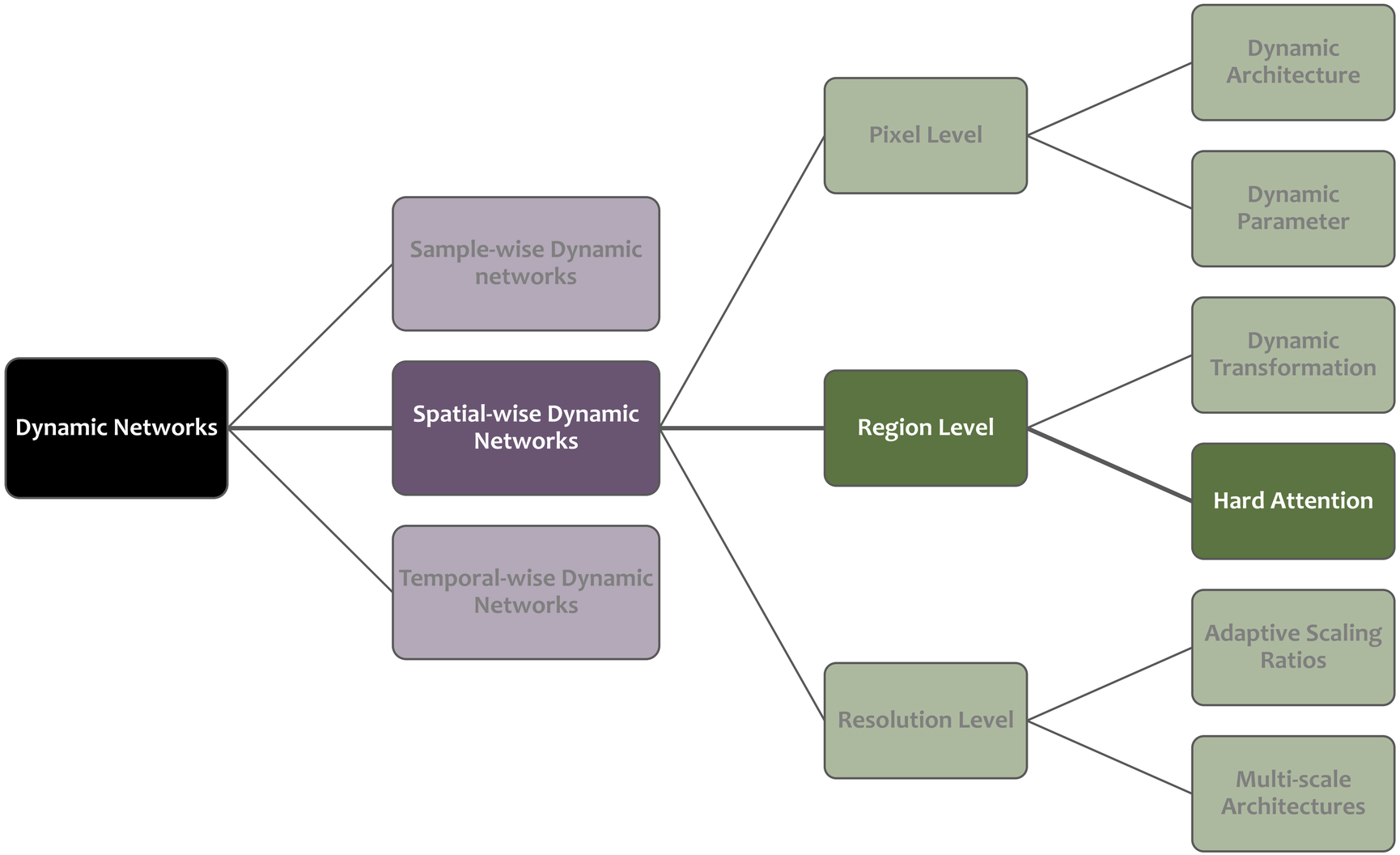
Pixel-level Dynamic Network

- **Sampling module:** generate a sampling indicator mask
- **Sparse Convolution:** compute features at sampled points
- **Interpolation module:** reconstruct entire feature map

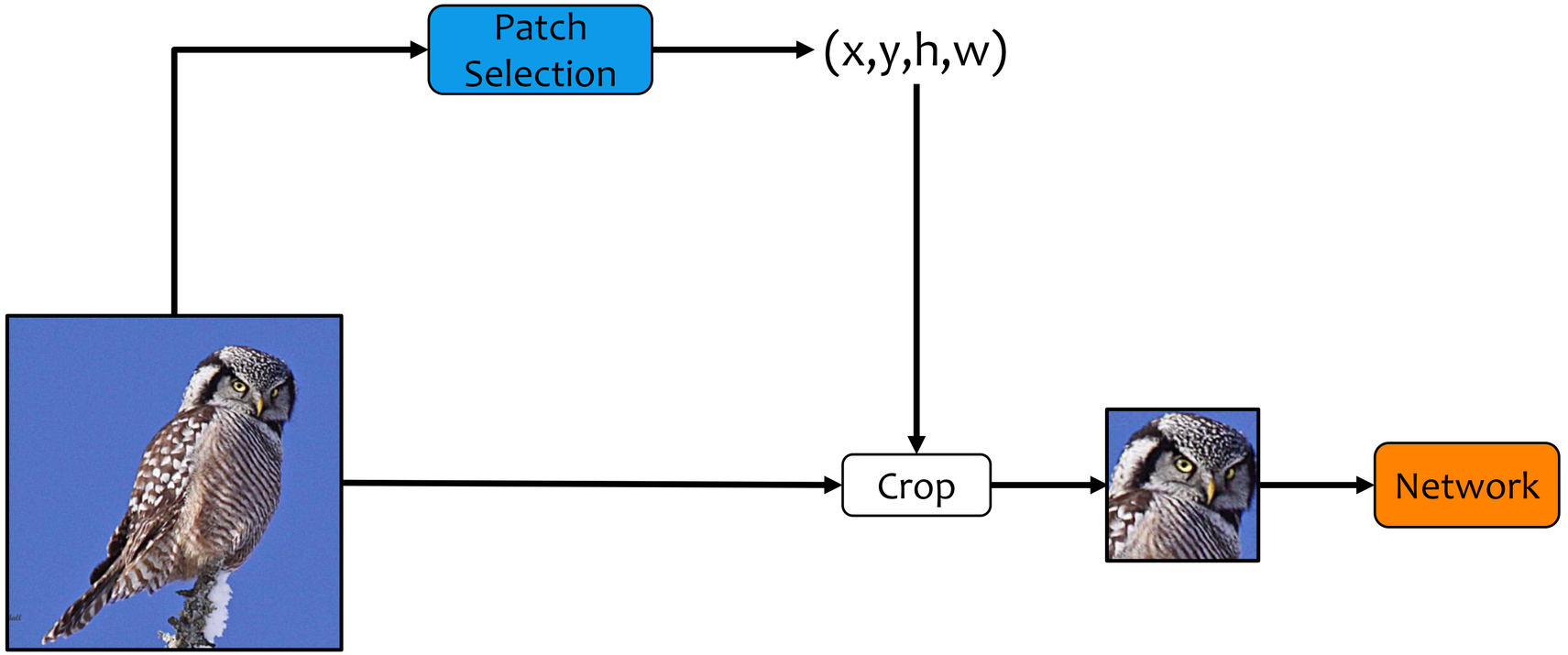


- Xie, Z., Zhang, Z., Zhu, X., Huang, G., & Lin, S. (2020, August). Spatially adaptive inference with stochastic feature sampling and interpolation. ECCV Oral.

Spatial-wise Dynamic Neural Networks

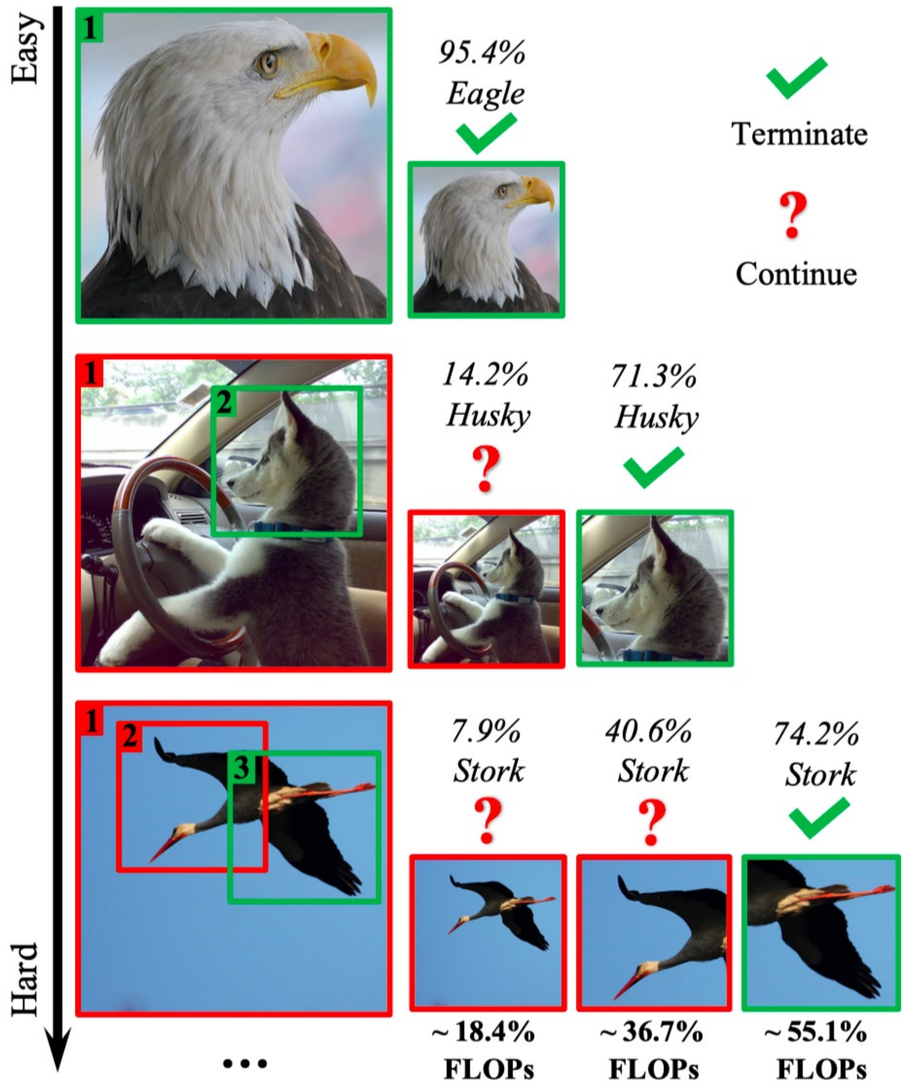


Region-level Dynamic Network



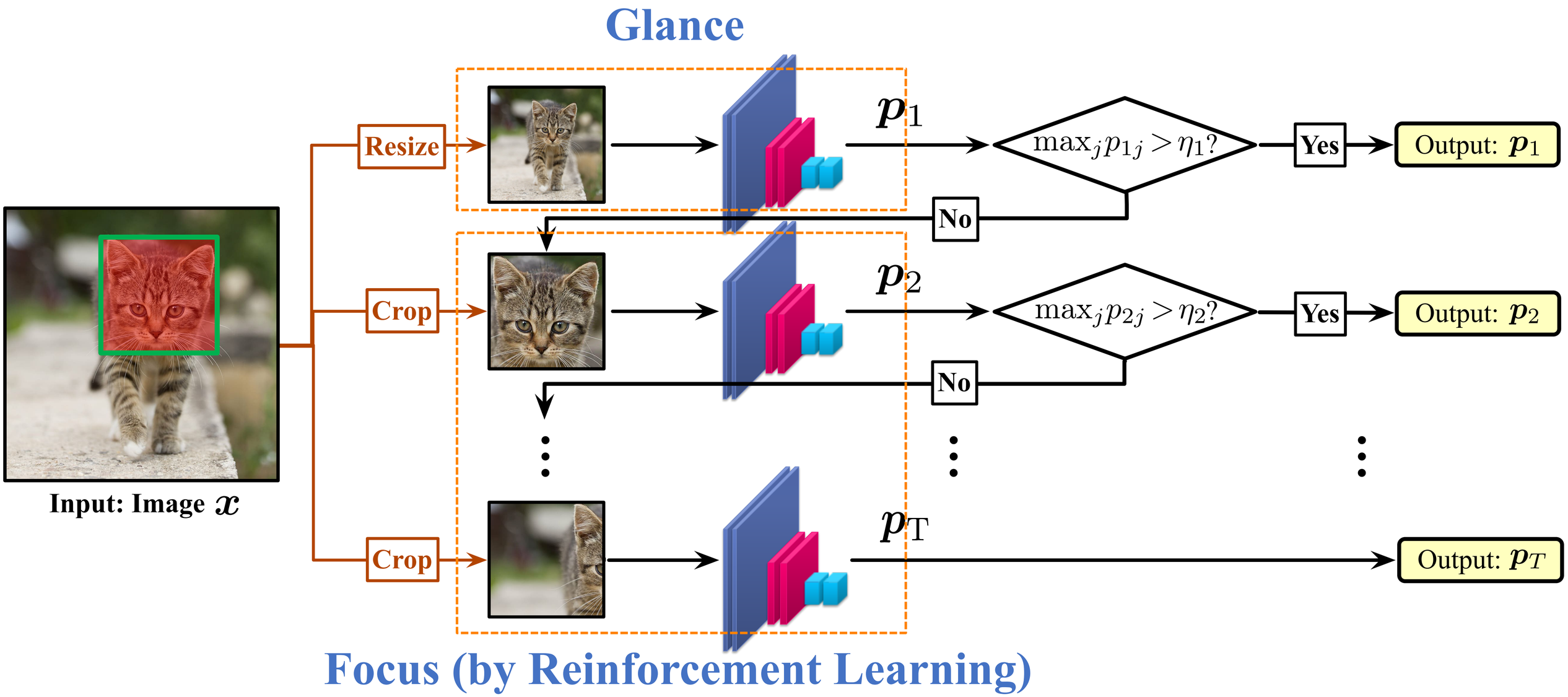
Region-level Dynamic Network

Human visual system processes information progressively.

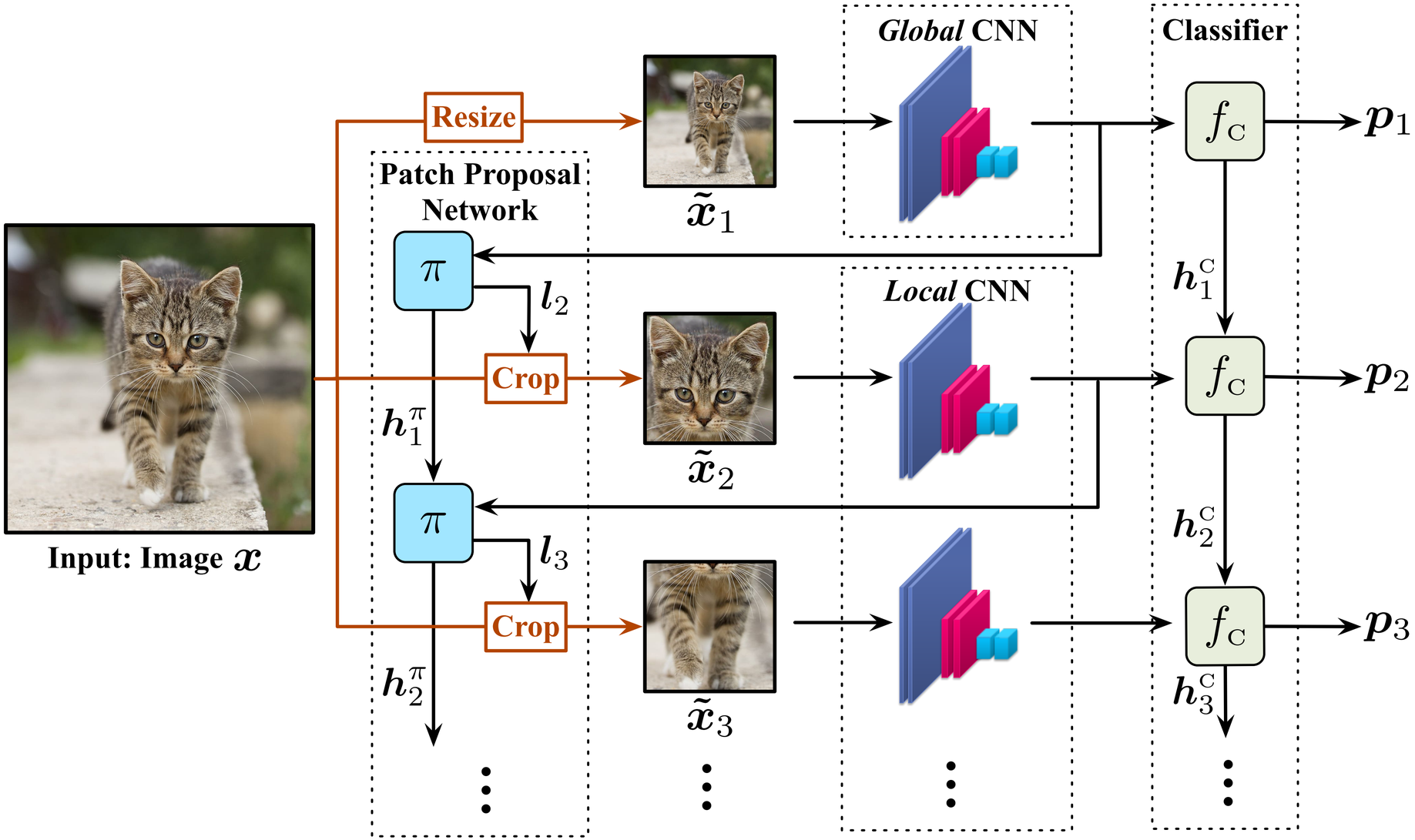


- Wang, Y., Lv, K., Huang, R., Song, S., Yang, L., & Huang, G. (2020). Glance and focus: a dynamic approach to reducing spatial redundancy in image classification. arXiv preprint arXiv:2010.05300.

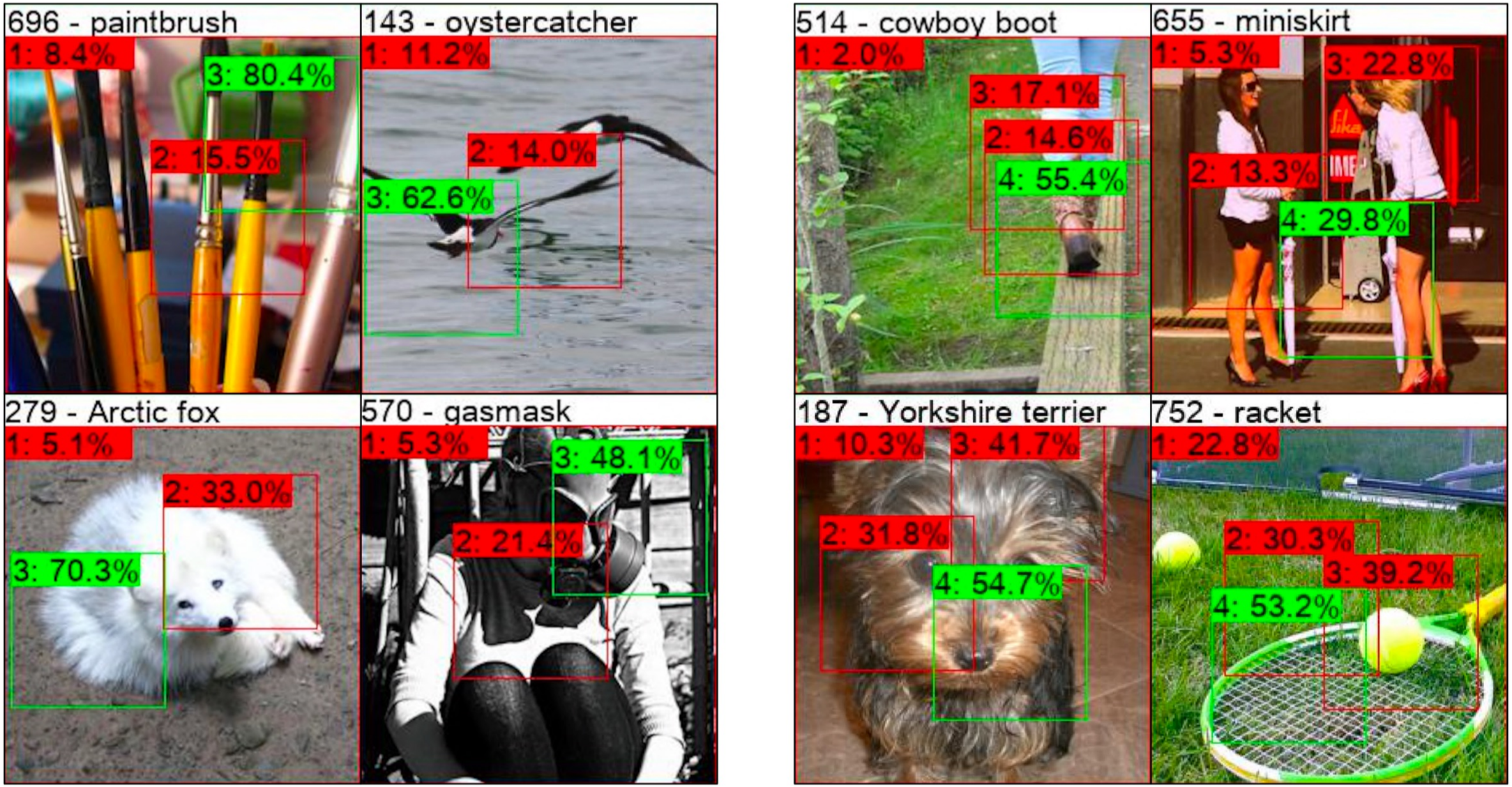
A Two-stage Framework



Network Architecture

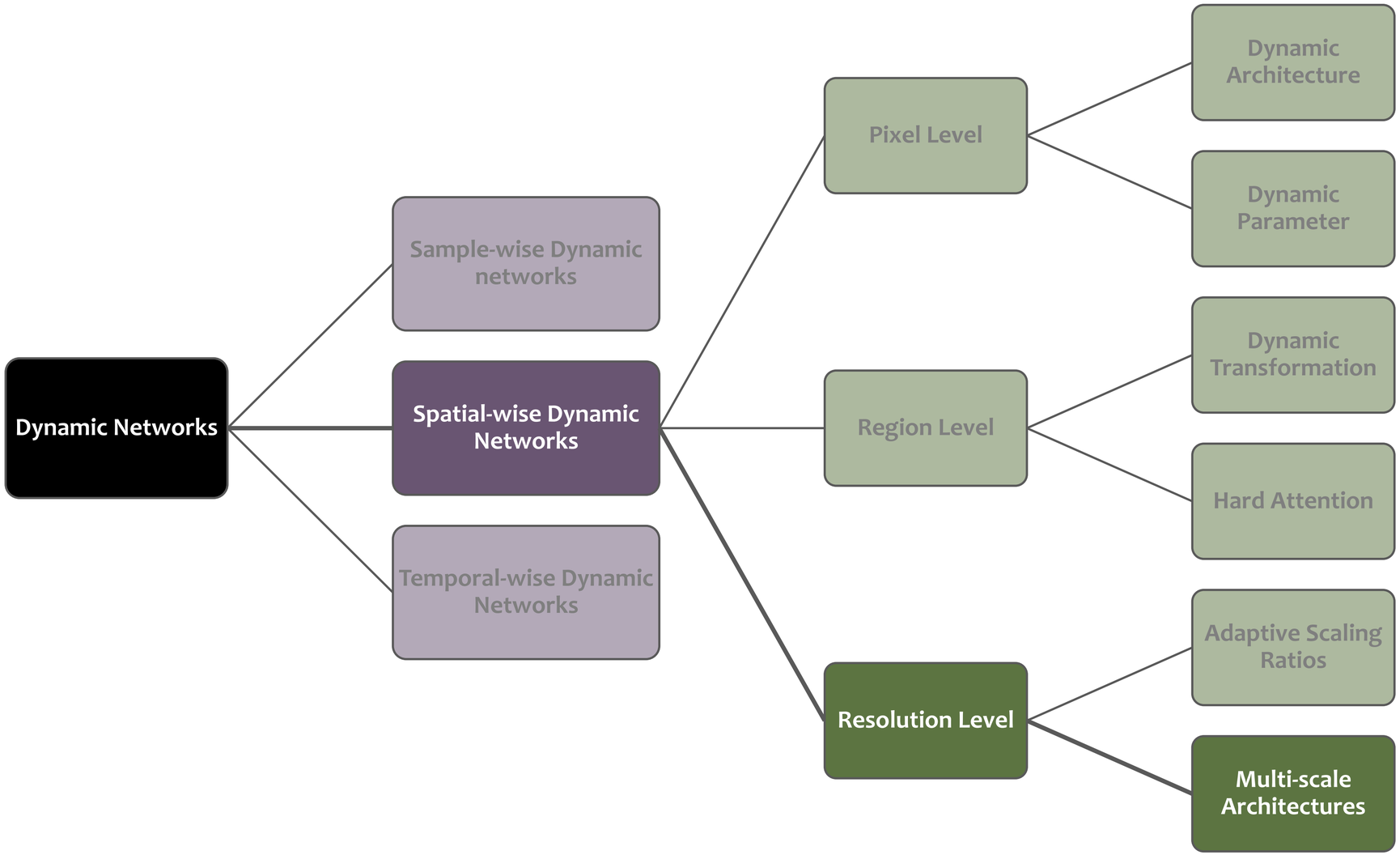


Visualization



77

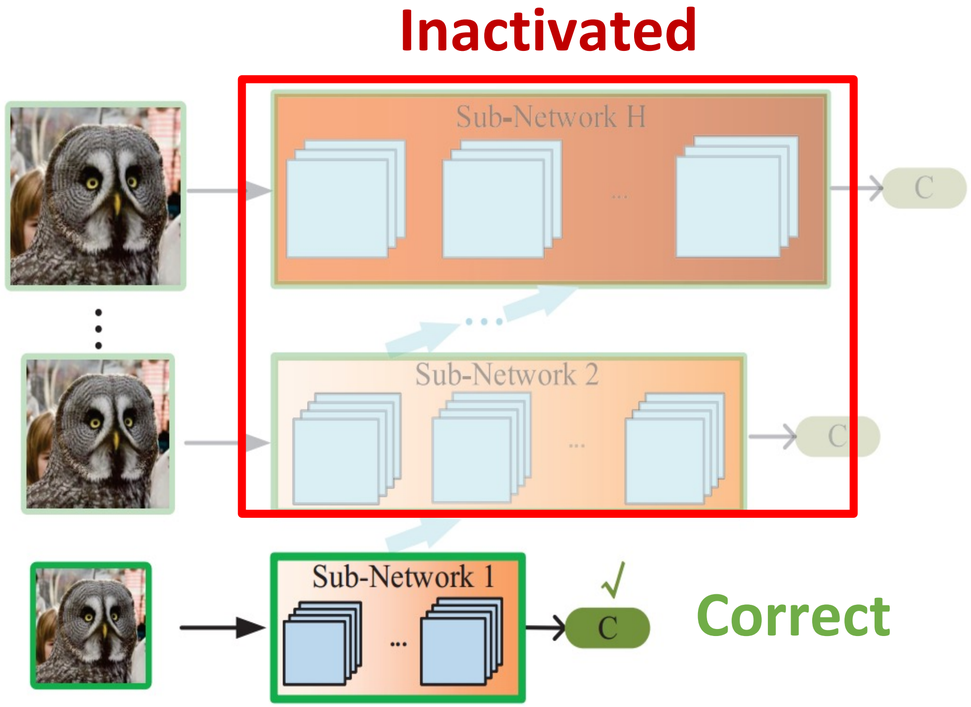
Spatial-wise Dynamic Neural Networks



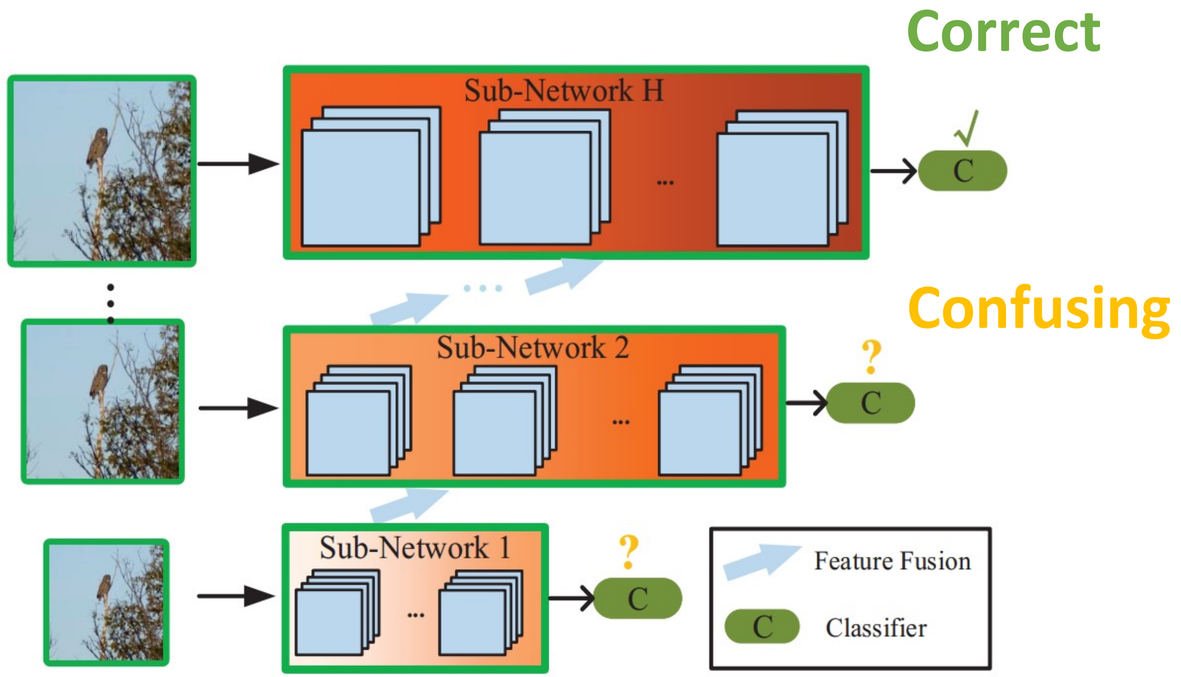
Low-resolution *representations are sufficient to recognize* ***"easy"*** *samples.*

Resolution Adaptation

- **Easy samples** (e.g. images containing large objects):

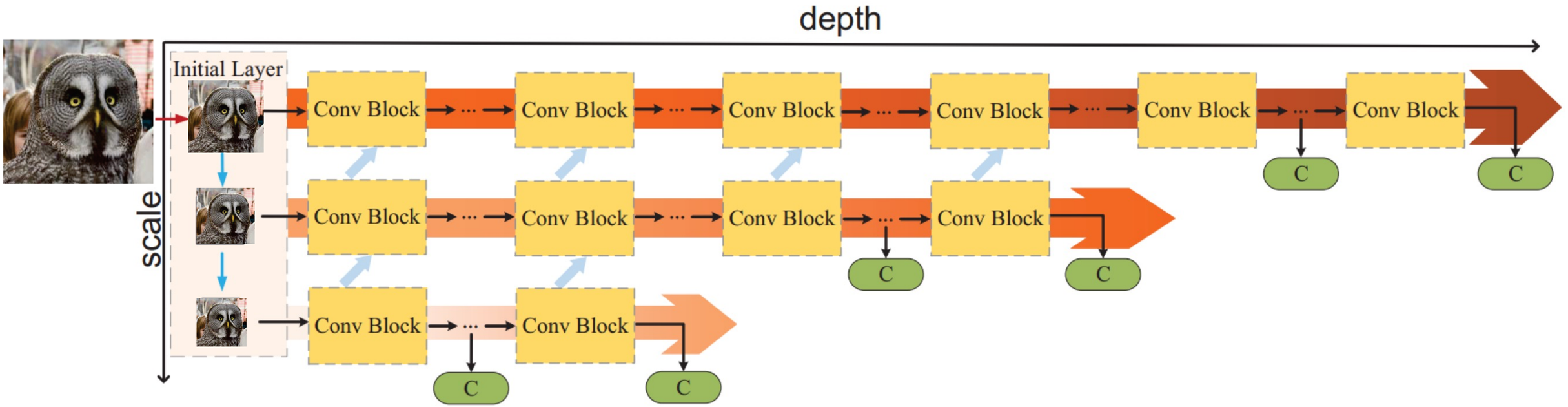


- **Hard samples** (e.g. images containing tiny objects) :



- Yang, L., Han, Y., Chen, X., Song, S., Dai, J., & Huang, G. (2020). Resolution adaptive networks for efficient inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2369-2378).

Resolution Adaptive Network



(Prediction Confidence (P_1) < ε_1)

Confusing

(Prediction Confidence > ε_2)

Exit

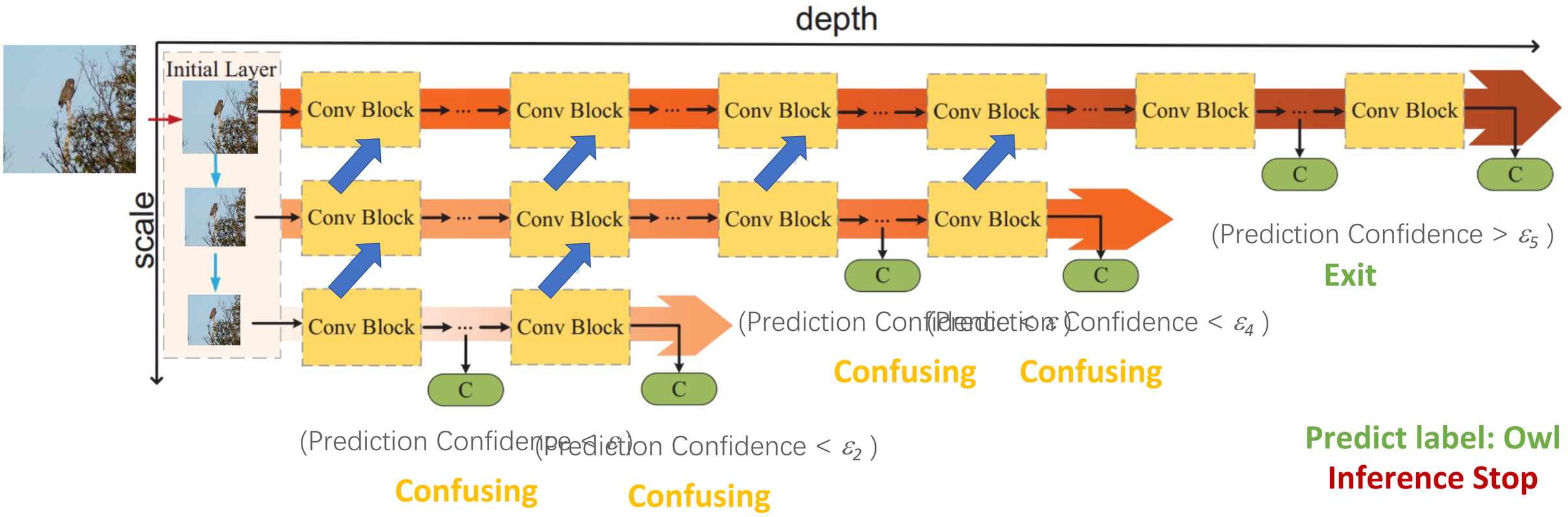
Predict label: Owl

Inference Stop

- Threshold (ε_i) controlling exiting of a sample.

- Yang, L., Han, Y., Chen, X., Song, S., Dai, J., & Huang, G. (2020). Resolution adaptive networks for efficient inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2369-2378).

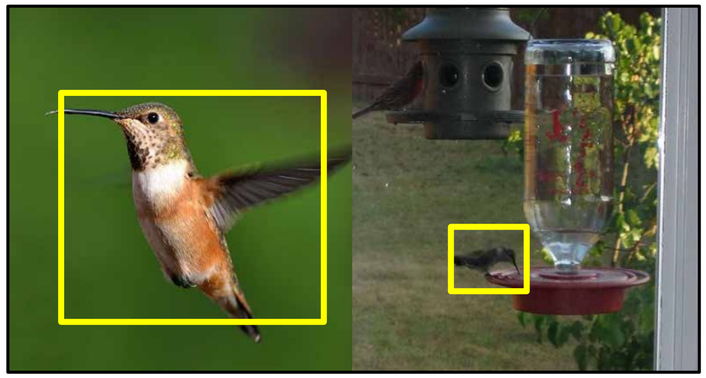
Resolution Adaptive Network



- Yang, L., Han, Y., Chen, X., Song, S., Dai, J., & Huang, G. (2020). Resolution adaptive networks for efficient inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2369-2378).

Visualization

Easy hard

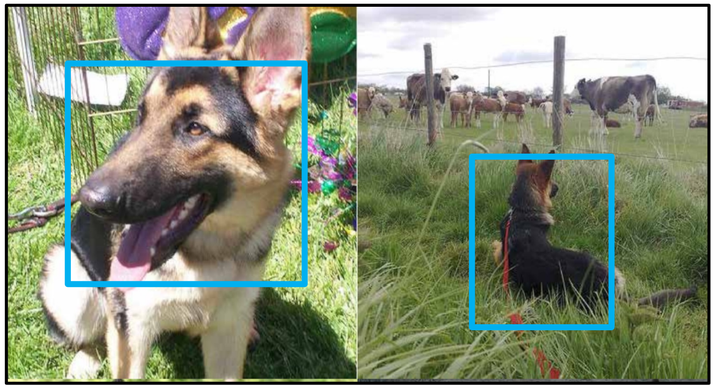


■ Images with tiny objects can be hard samples.

Easy Hard

■ Images with multiple objects can be hard samples.

Easy Hard

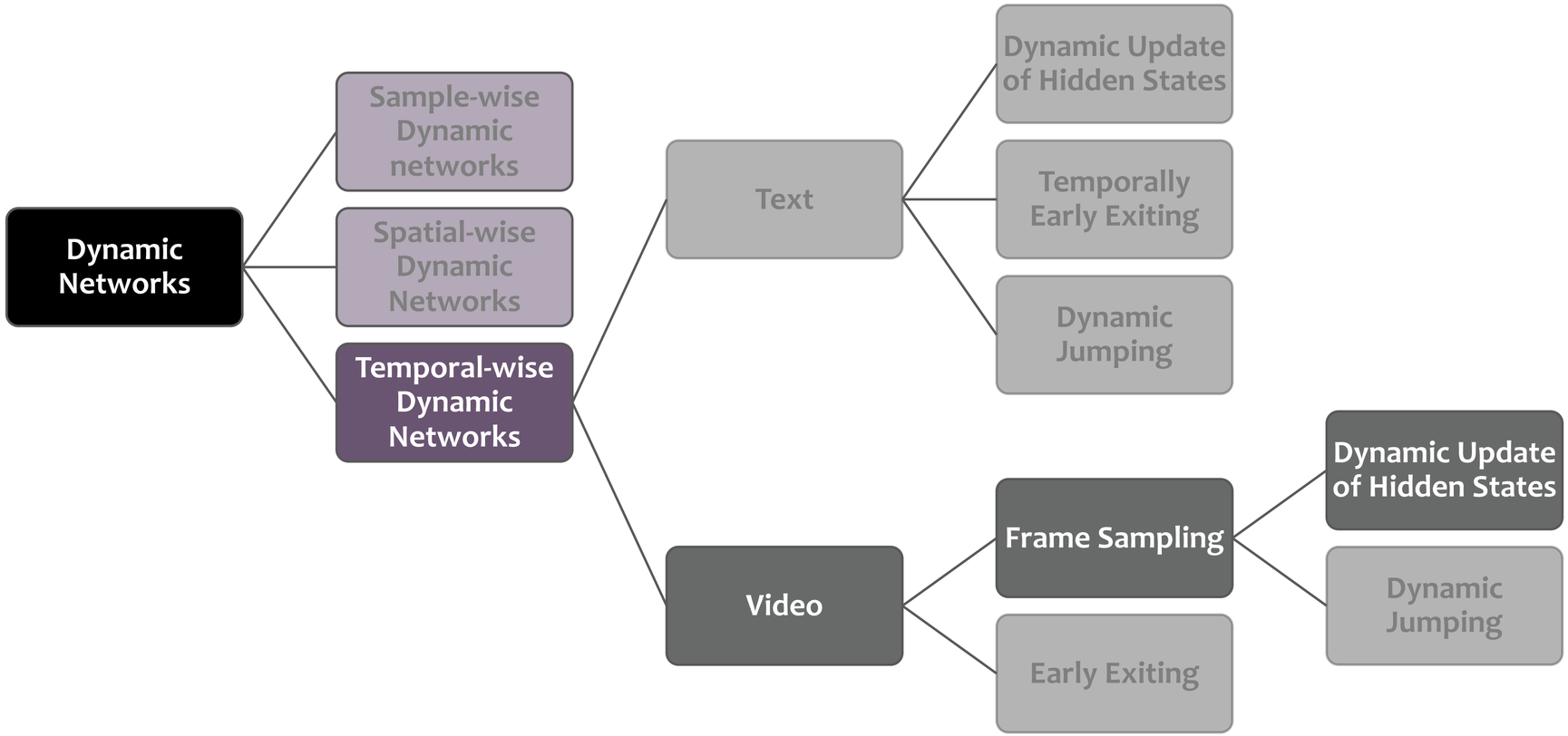


■ Images with objects w/o representative characteristics can be hard samples.

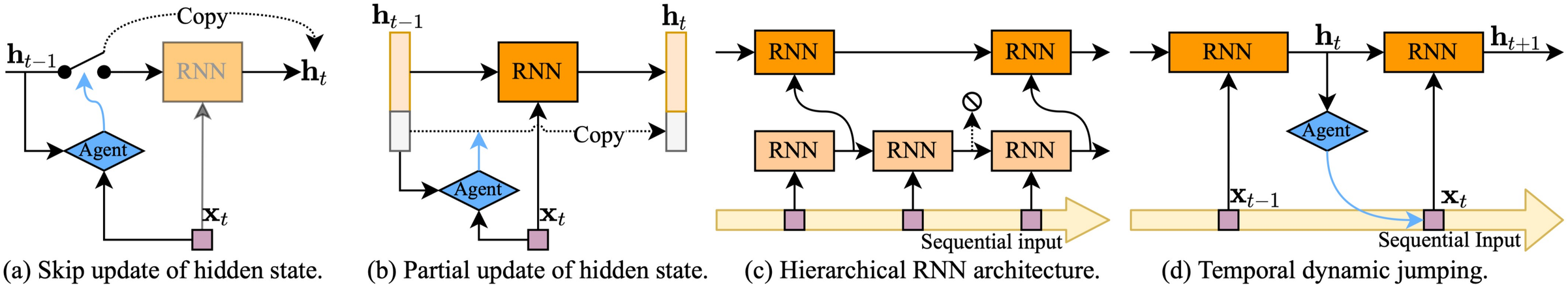
Overview

- Introduction
- Sample-wise Dynamic Networks
- Spatial-wise Dynamic Networks
- **Temporal-wise Dynamic Networks**
- Inference & Training
- Applications
- Discussion

Dynamic Architecture



RNN-based Approaches

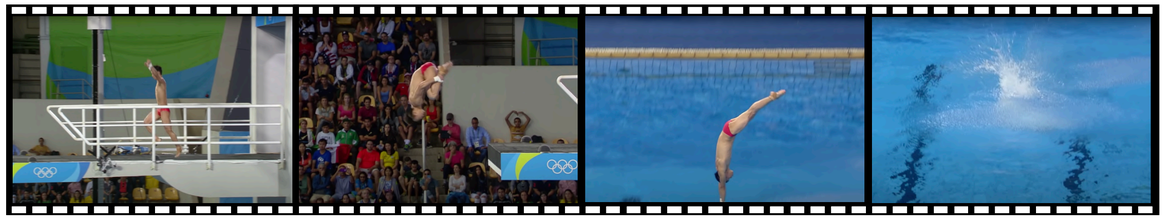


(a) Skip update of hidden state. (b) Partial update of hidden state. (c) Hierarchical RNN architecture. (d) Temporal dynamic jumping.

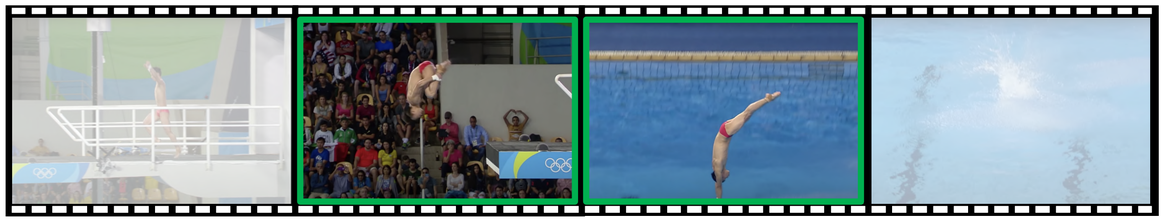
- Campos, V., Jou, B., Giró-i-Nieto, X., Torres, J., & Chang, S. F. (2017). Skip rnn: Learning to skip state updates in recurrent neural networks. arXiv preprint arXiv:1708.06834.
- Seo, M., Min, S., Farhadi, A., & Hajishirzi, H. (2017). Neural speed reading via skim-rnn. arXiv preprint arXiv:1711.02085.
- Junyoung Chung, Sungjin Ahn, and Yoshua Bengio. Hierarchical multiscale recurrent neural networks. In ICLR, 2017.
- Adams Wei Yu, Hongrae Lee, and Quoc Le. Learning to Skim Text. In ACL, 2017.

Adaptive Focus for Efficient Video Recognition

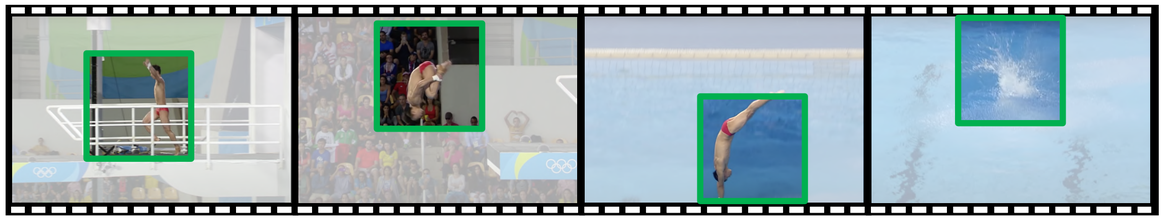
Patch-level spatial-wise +
Temporal-wise (frame skipping)



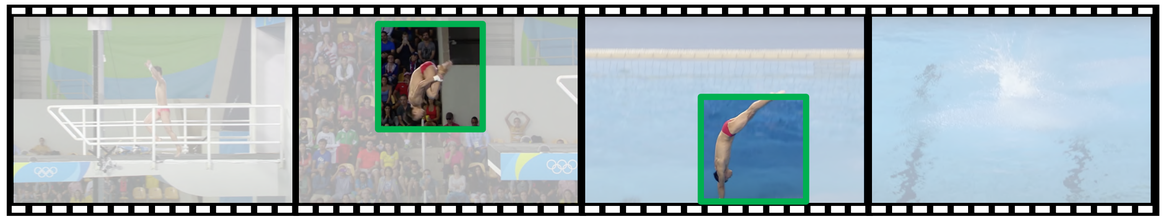
(a) Input Video (label: diving)



(b) Temporal-based Methods (existing works)

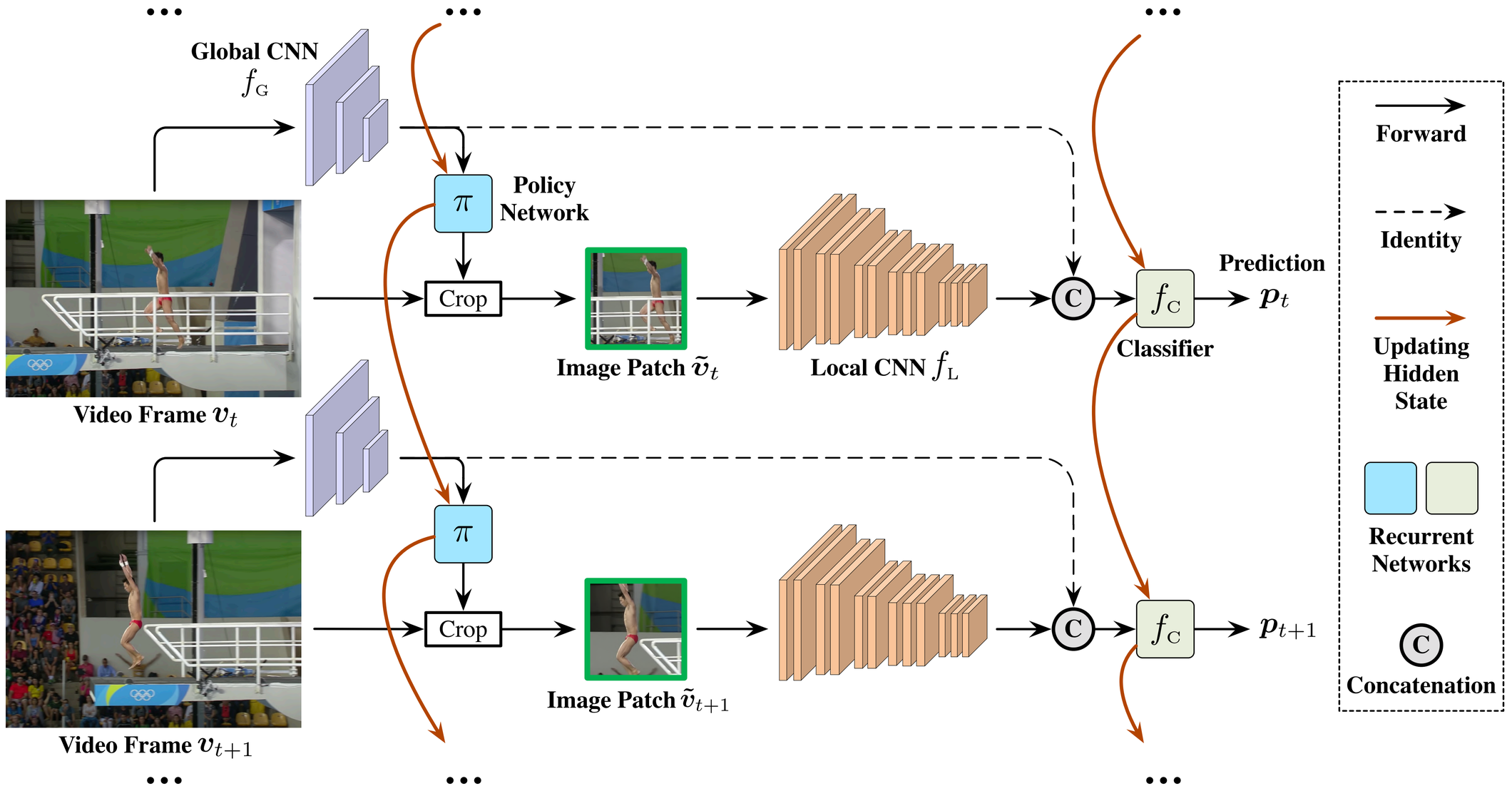


(c) AdaFocus (ours)



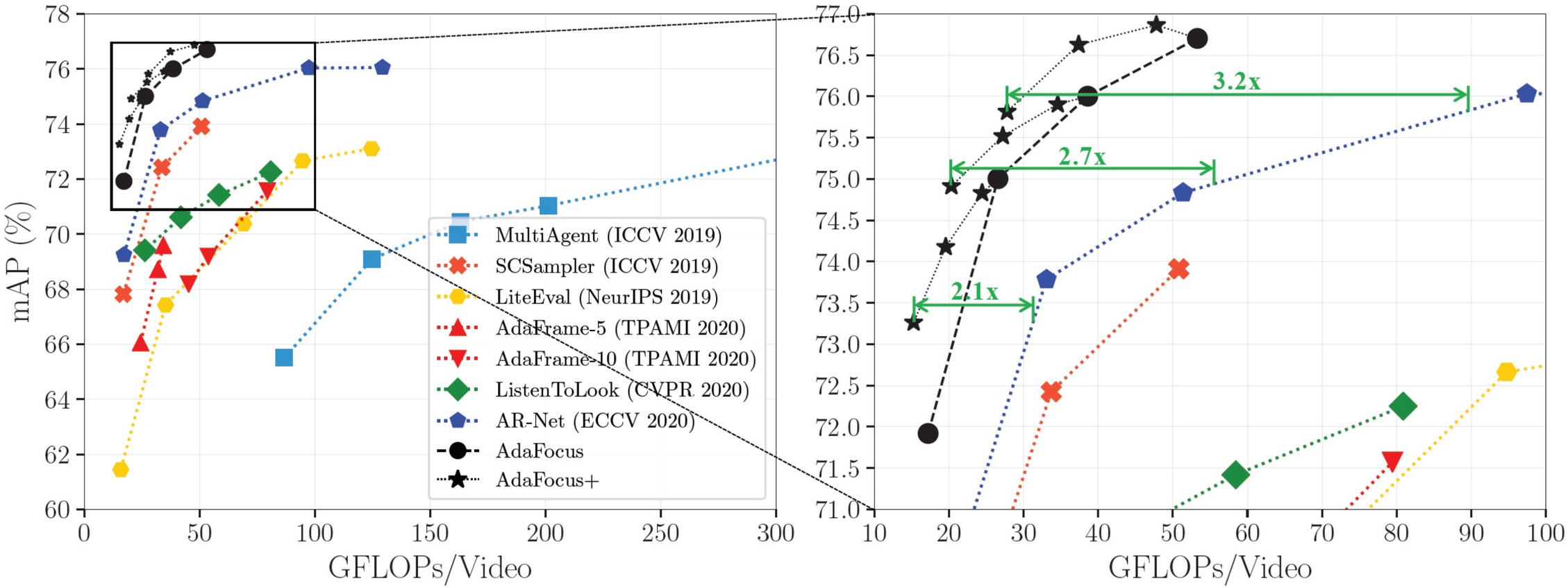
(d) AdaFocus+

- Wang, Y., Chen, Z., Jiang, H., Song, S., Han, Y., & Huang, G. (2021). Adaptive Focus for Efficient Video Recognition. In *ICCV*.



- Wang, Y., Chen, Z., Jiang, H., Song, S., Han, Y., & Huang, G. (2021). Adaptive Focus for Efficient Video Recognition. In *ICCV*.

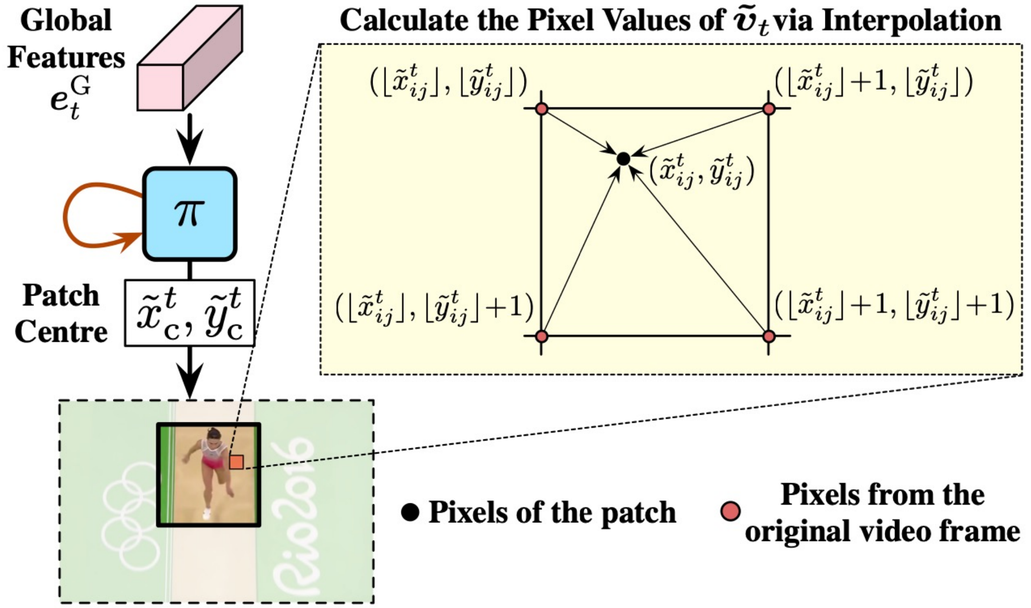
Offline Video Recognition on ActivityNet

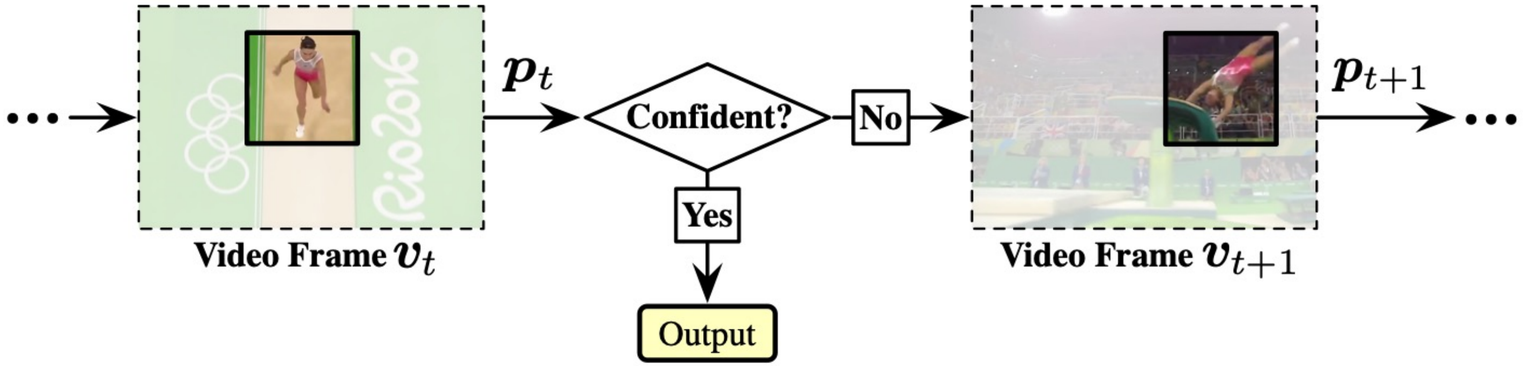


- Wang, Y., Chen, Z., Jiang, H., Song, S., Han, Y., & Huang, G. (2021). Adaptive Focus for Efficient Video Recognition. In *ICCV*.

AdaFocus-v2: end-to-end training

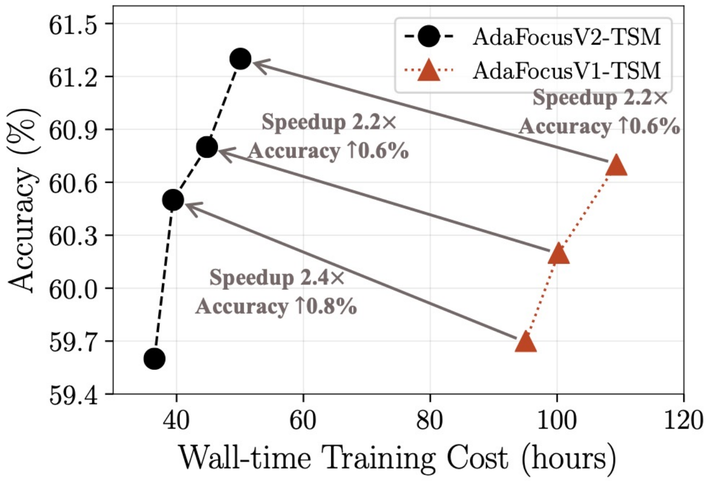
	AdaFocusV1	AdaFocusV2
Pre-training	❶ Pre-train f_{G} on Sth-Sth V1. ❷ Pre-train f_{L} on Sth-Sth V1.	***End-to-End Training*** $(f_{\mathrm{G}}, f_{\mathrm{L}}, f_{\mathrm{C}}, \pi)$
Stage-1	❸ Train f_{L} and f_{C} using random patches.	
Stage-2	❹ Train π using reinforcement learning.	
Stage-3	❺ Fine-tune f_{L} and f_{C}.	

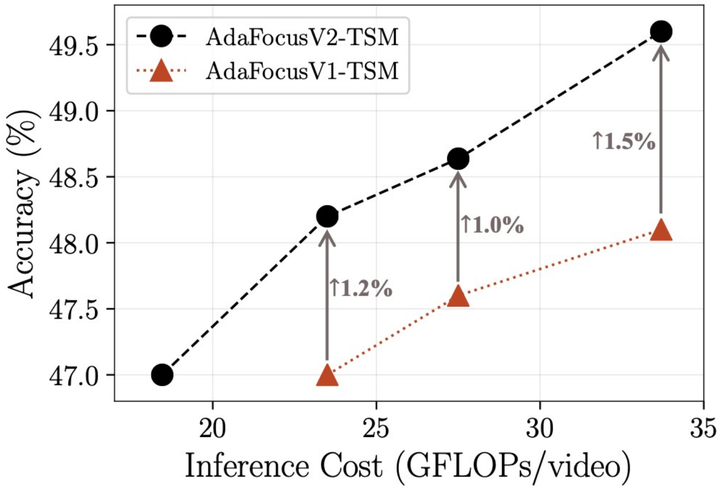


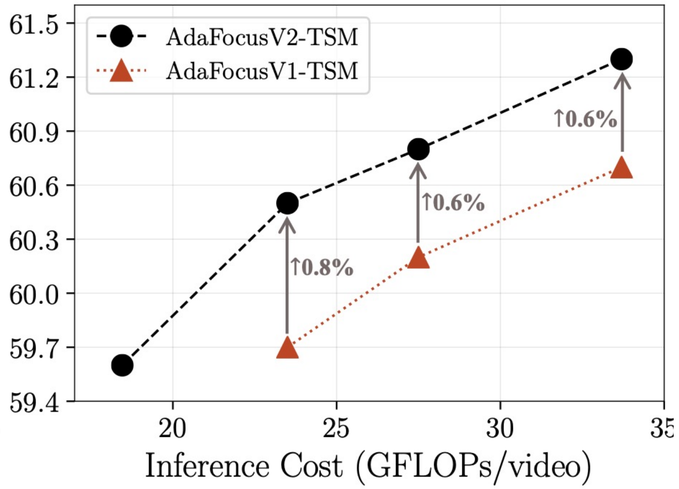


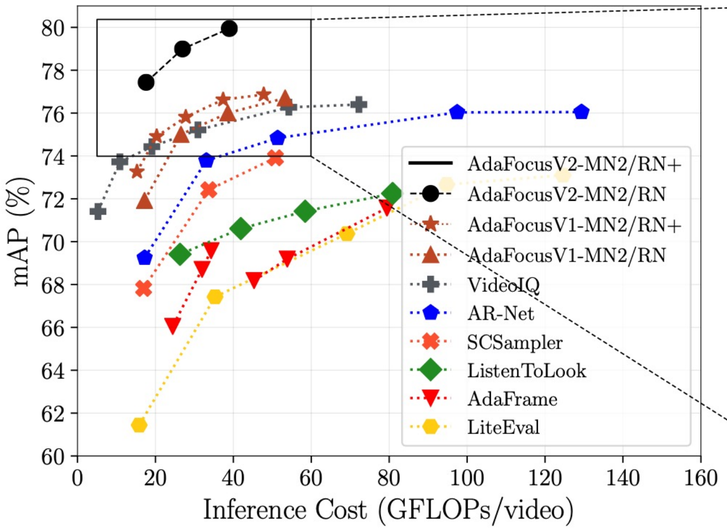
- Wang, Y., Yue, Y., Lin, Y., Jiang, H., Lai, Z., Kulikov, V., ... & Huang, G. (2021). AdaFocus V2: End-to-End Training of Spatial Dynamic Networks for Video Recognition. arXiv preprint arXiv:2112.14238.

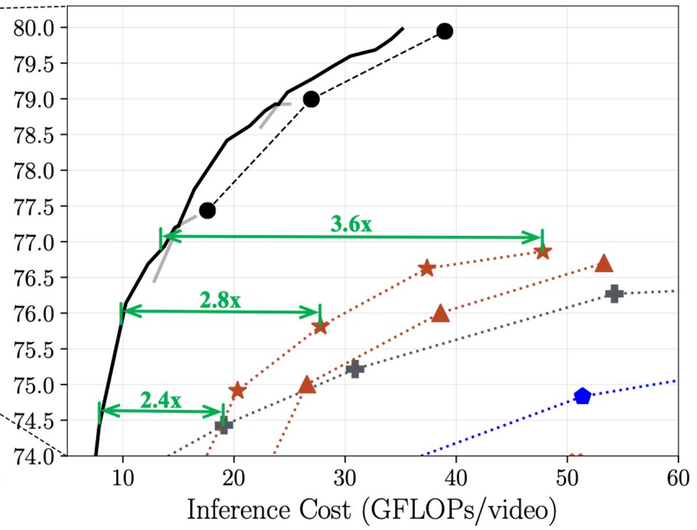
AdaFocus-v2: end-to-end training

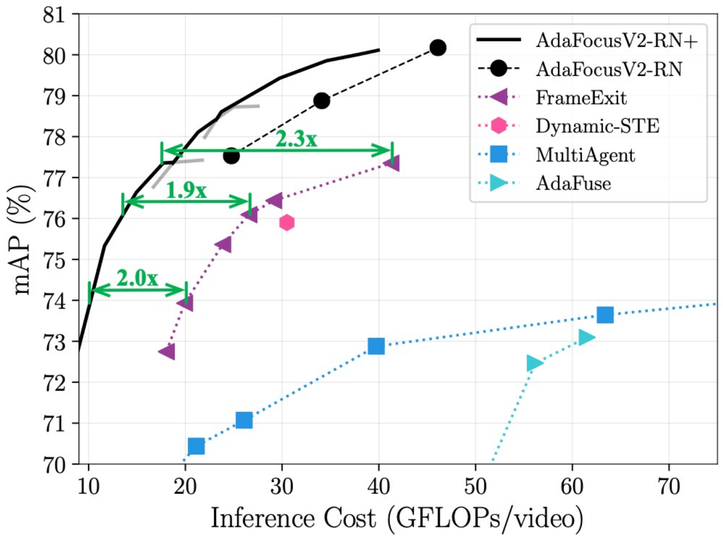












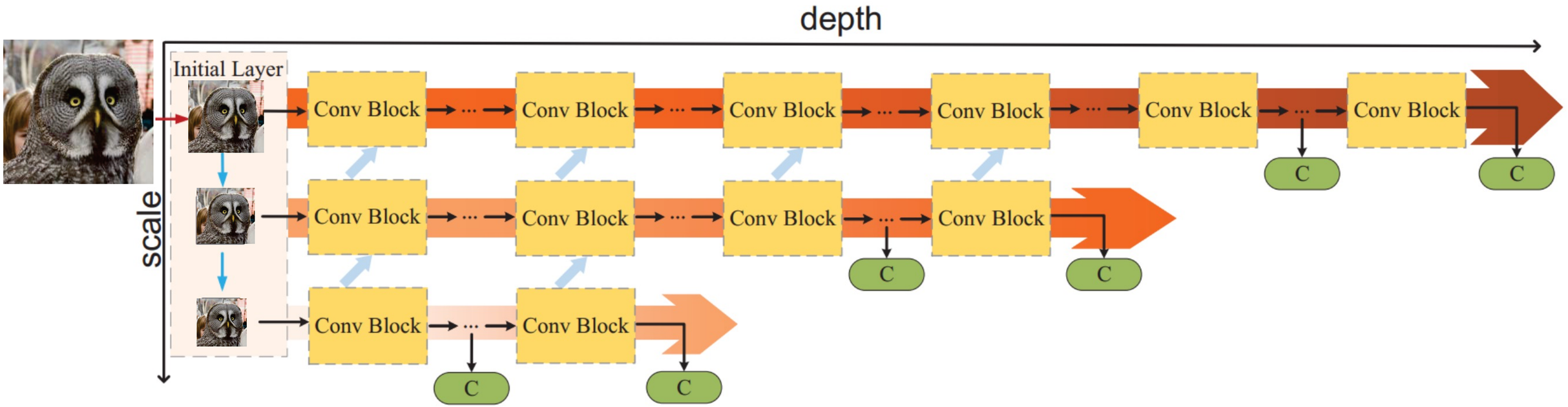
- Wang, Y., Yue, Y., Lin, Y., Jiang, H., Lai, Z., Kulikov, V., ... & Huang, G. (2021). AdaFocus V2: End-to-End Training of Spatial Dynamic Networks for Video Recognition. arXiv preprint arXiv:2112.14238.

Overview

清華大學
Tsinghua University

- Introduction
- Sample-wise Dynamic Networks
- Spatial-wise Dynamic Networks
- Temporal-wise Dynamic Networks
- **Inference & Training**
- Applications
- Discussion

Decision Making: Based on Confidence



(Prediction Confidence (P_1) < ε_1)

Confusing

(Prediction Confidence > ε_2)

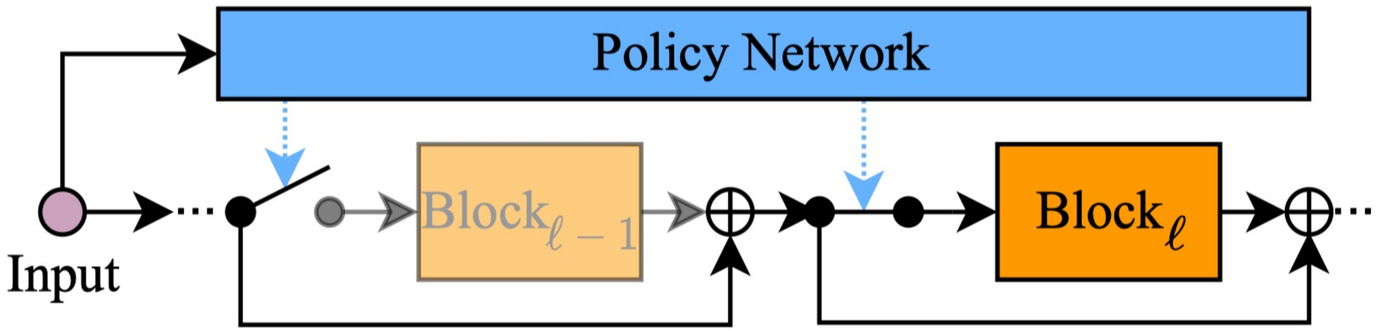
Exit

Predict label: Owl

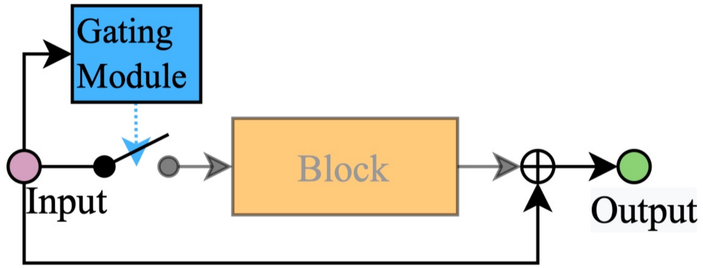
Inference Stop

- Threshold (ε_i) controlling exiting of a sample.

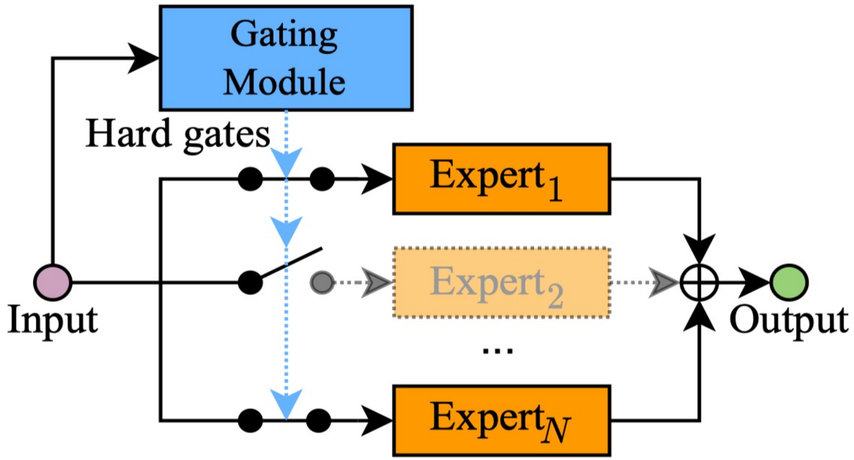
Decision Making: Based on Policy Networks



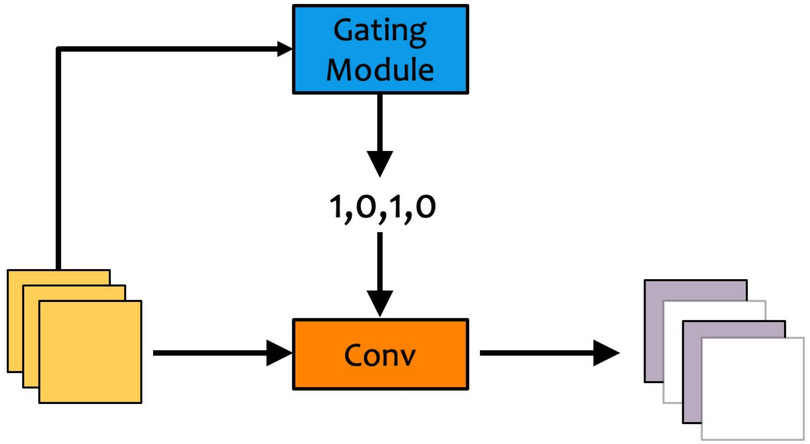
Decision Making: Based on Gating Functions



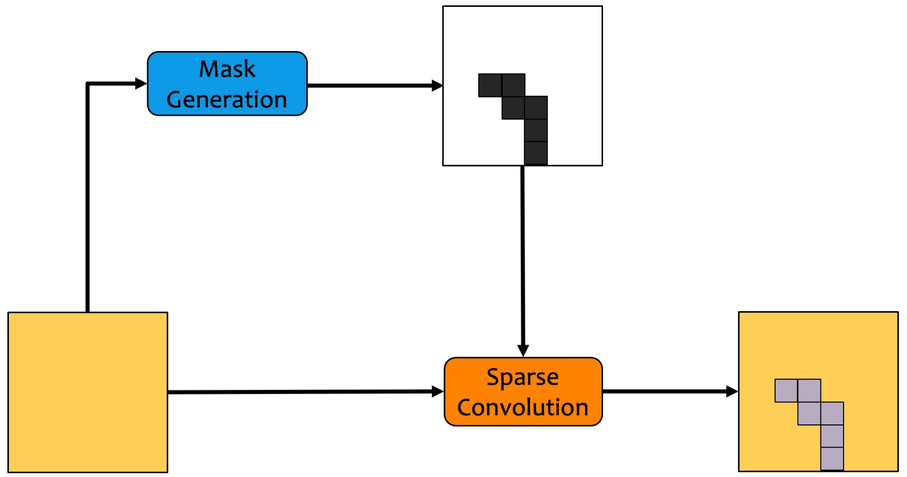
Layer Skipping



Branch skipping



Channel skipping



Spatial-wise dynamic convolution

Optimization for Discrete Decisions

Tsinghua University

Gradient Estimation

- Gumbel SoftMax, for training plug-in modules
  - Layer skipping
  - Channel skipping
  - Pixel selection
  - …

Reinforcement Learning

- For training policy networks
  - Layer skipping
  - Patch selection
  - Frame selection
  - …

Training Objectives for Encouraging Sparsity

$$L = L_{\text{task}} + L_{\text{efficiency}}$$

- Activation rate of layers/channels/selected locations
- FLOPs

Overview

- Introduction
- Sample-wise Dynamic Networks
- Spatial-wise Dynamic Networks
- Temporal-wise Dynamic Networks
- Inference & Training
- **Applications**
- Discussion

Application of Dynamic Networks

Tsinghua University

Fields	Data	Type	Subfields & references
Computer Vision	Image	Sa	Object detection (face [40], [203], [204], facial point [205], pedestrian [206], general [33], [207], [208], [209], [210]) Image segmentation [107], [211], [212], Super resolution [213], Style transfer [214], Coarse-to-fine classification [215]
		Sa & Sp	Image segmentation [34], [129], [146], [148], [150], [154], [156], [216], [217], [218], [219], [220], Image-to-image translation [221], Object detection [110], [111], [147], [148], [164], Semantic image synthesis [222], [223], [224], Image denoising [225], Fine-grained classification [158], [162], [226], [227] Eye tracking [158], Super resolution [151], [153], [228]
		Sa & Sp & Te	General classification [39], [159], [161], Multi-object classification [229], [230], Fine-grained classification [160]
	Video	Sa	Multi-task learning (human action recognition and frame prediction) [231]
		Sa & Te	Classification (action recognition) [61], [177], [181], [189], [190], [191], [192], [196], [232], Semantic segmentation [233] Video face recognition [22], [188], Action detection [179], [180], Action spotting [178], [187]
		Sa & Sp & Te	Classification [196], [197], Frame interpolation [234], [235], Super resolution [236], Video deblurring [237], [238], Action prediction [239]
	Point Cloud	Sa & Sp	3D Shape classification and segmentation, 3D scene segmentation [240], 3D semantic scene completion [241]
Natural Language Processing	Text	Sa	Neural language inference, Text classification, Paraphrase similarity matching, and Sentiment analysis [59], [60]
		Sa & Te	Language modeling [11], [16], [118], [170], [172], Machine translation [16], [35], [36], Classification [64], [65], [174], Sentiment analysis [168], [169], [171], [175], [176], Question answering [35], [63], [168], [171], [173]
Cross-Field	Image captioning [130], [242], Video captioning [243], [244], Visual question answering [123], [124], [245], Multi-modal sentiment analysis [246], [247]		
Others	Time series forecasting [248], [249], [250], Link prediction [251], Recommendation system [77], [252], [253], [254] Graph classification [121], Document classification [156], [255], [256], [257], Stereo confidence estimation [258]		

Overview

清華大學
Tsinghua University

- Introduction
- Sample-wise Dynamic Networks
- Spatial-wise Dynamic Networks
- Temporal-wise Dynamic Networks
- Inference & Training
- Applications
- **Discussion**

Advantages

Efficiency

Representation Power

Adaptiveness

Compatibility

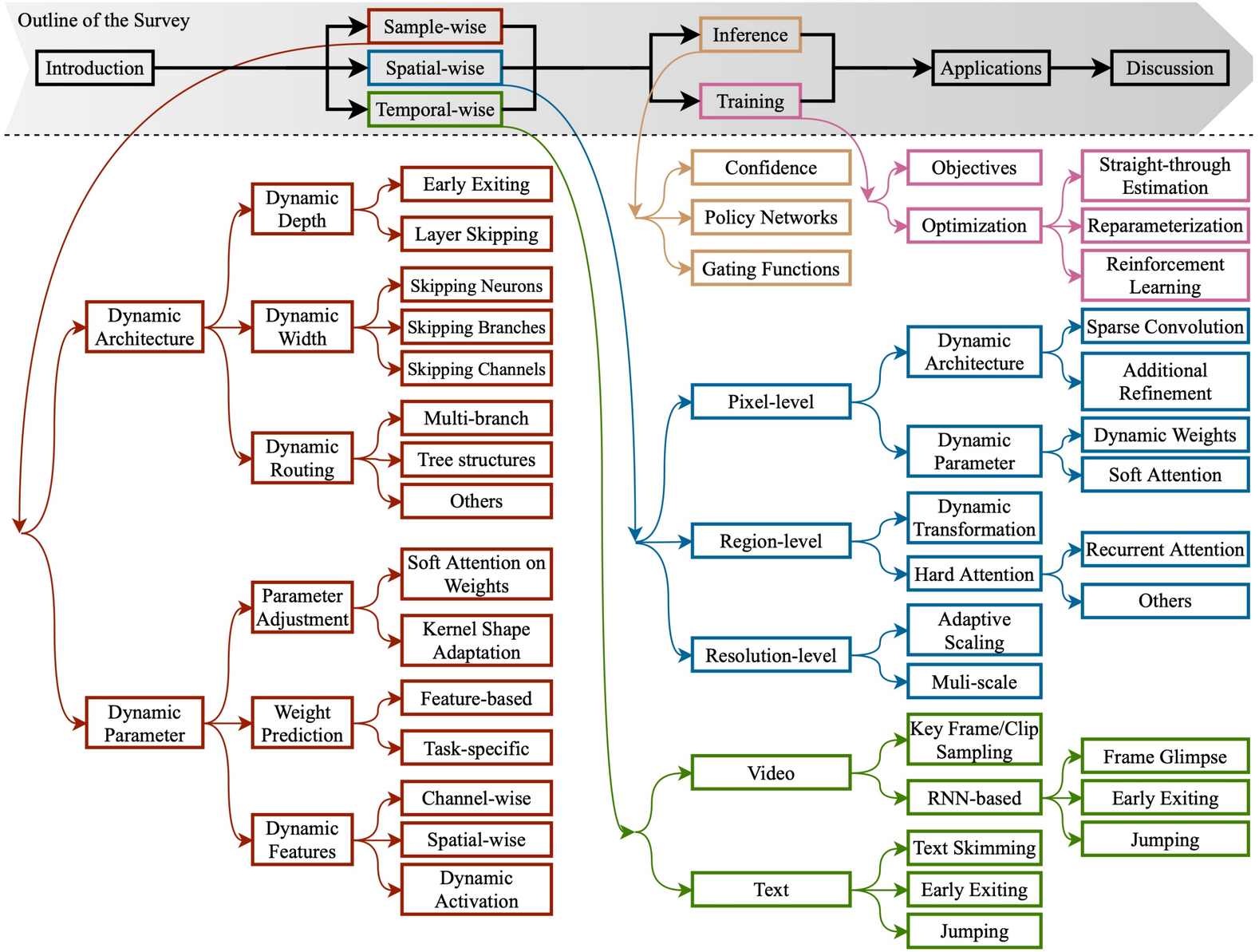
Generality

Interpretability

Challenges in Dynamic Neural Networks

Theories	Architecture Design	Applicability on more diverse tasks
Gap between theoretical & practical efficiency	Robustness	Interpretability

A (relatively) comprehensive **survey** on dynamic neural networks



arXiv Paper

Han & Huang, et al, Dynamic neural networks: A survey, TPAMI

Thank you!

清華大學
Tsinghua University