SP-Net: Slowly Progressing Dynamic Inference Networks

Huanyu Wang**, Wenhu Zhang**, Shihao Su, Hui Wang, Zhenwei Miao, Xin Zhan, and Xi Li*

College of Computer Science and Technology, Polytechnic Institute, Zhejiang University; Alibaba DAMO Academy, HangZhou, Zhejiang.

Introduction

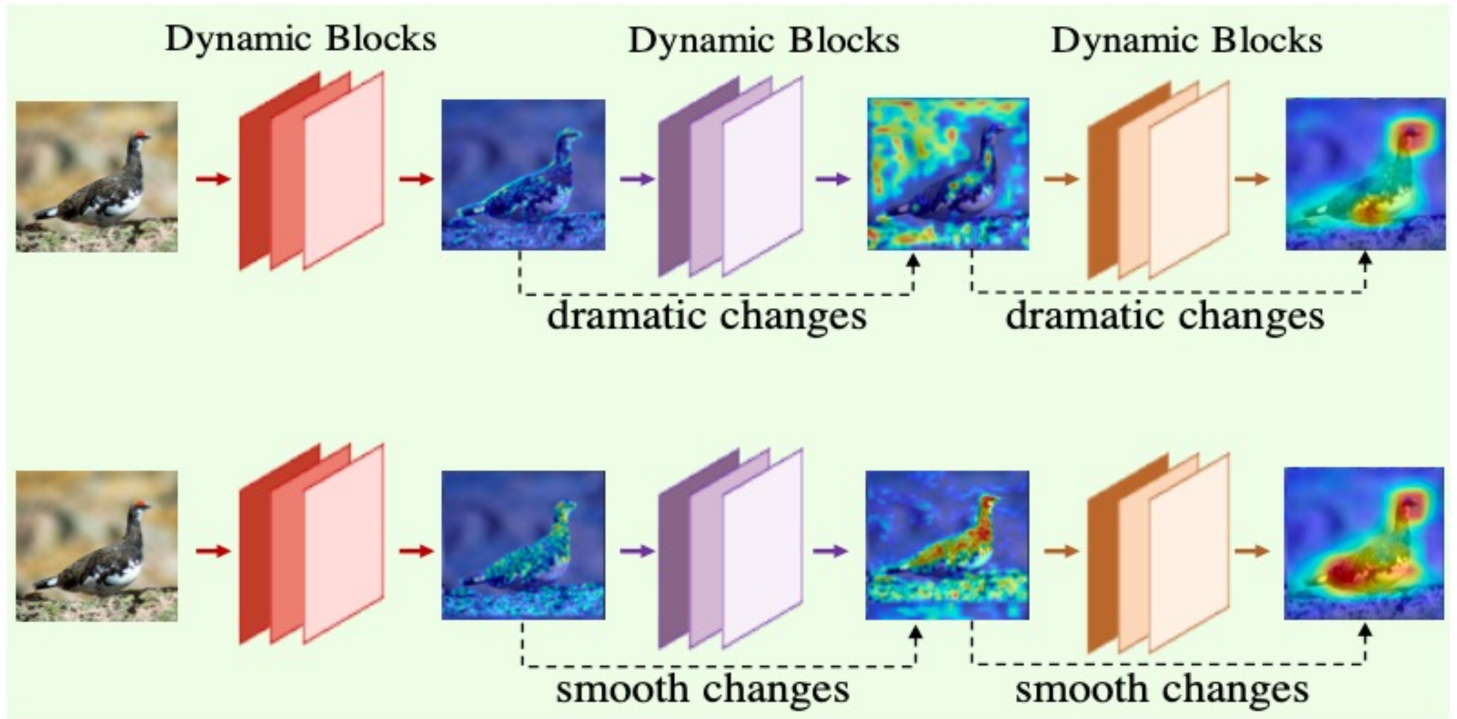
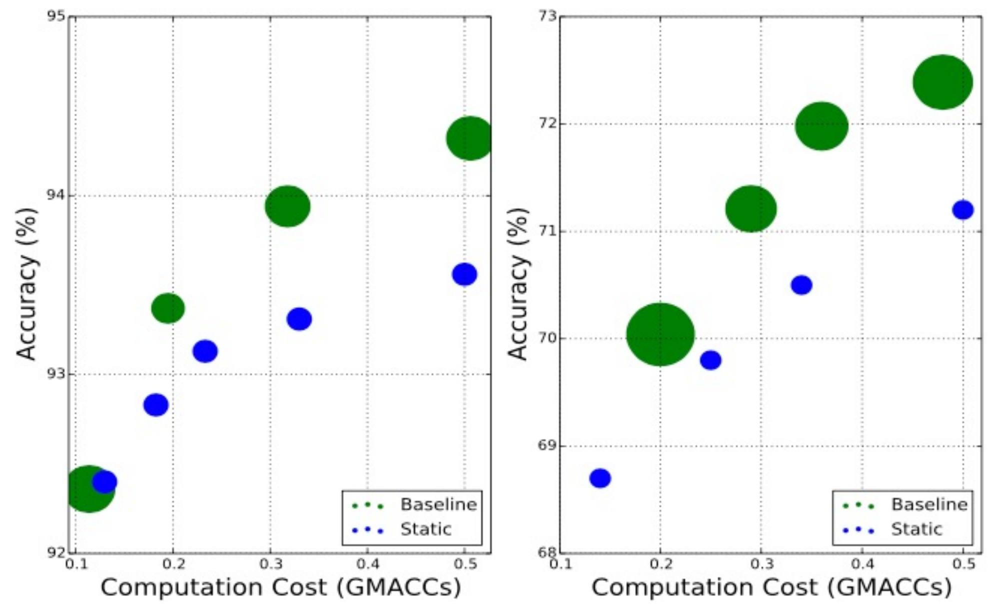


Illustration of features evolve across the dynamic inference network.

- Lower: The stable features evolution across the network.
- Upper: The unstable features evolution throughout the network.



- The stability of static networks (ResNets) and vanilla dynamic inference networks on CIFAR. The size of dots is the variance of results.
- The stability of static networks and vanilla dynamic inference networks on CIFAR-100. The size of dots is the variance of results.

Contribution

• We propose a slowly progressing dynamic inference network, which effectively stabilizes the optimization of dynamic inference networks.

• We slow down the progress in routers by taking advantage of the information from historical iterations to solve the unbalance of sensitivity between parameters in convolutional blocks and routers.

• We slow down the feature evolution by regularizing the direction of the feature evolution, making the feature evolves smoothly.

• We conduct experiments on three benchmarks and obtain state-of-the-art results in terms of performance and efficiency.

Framework

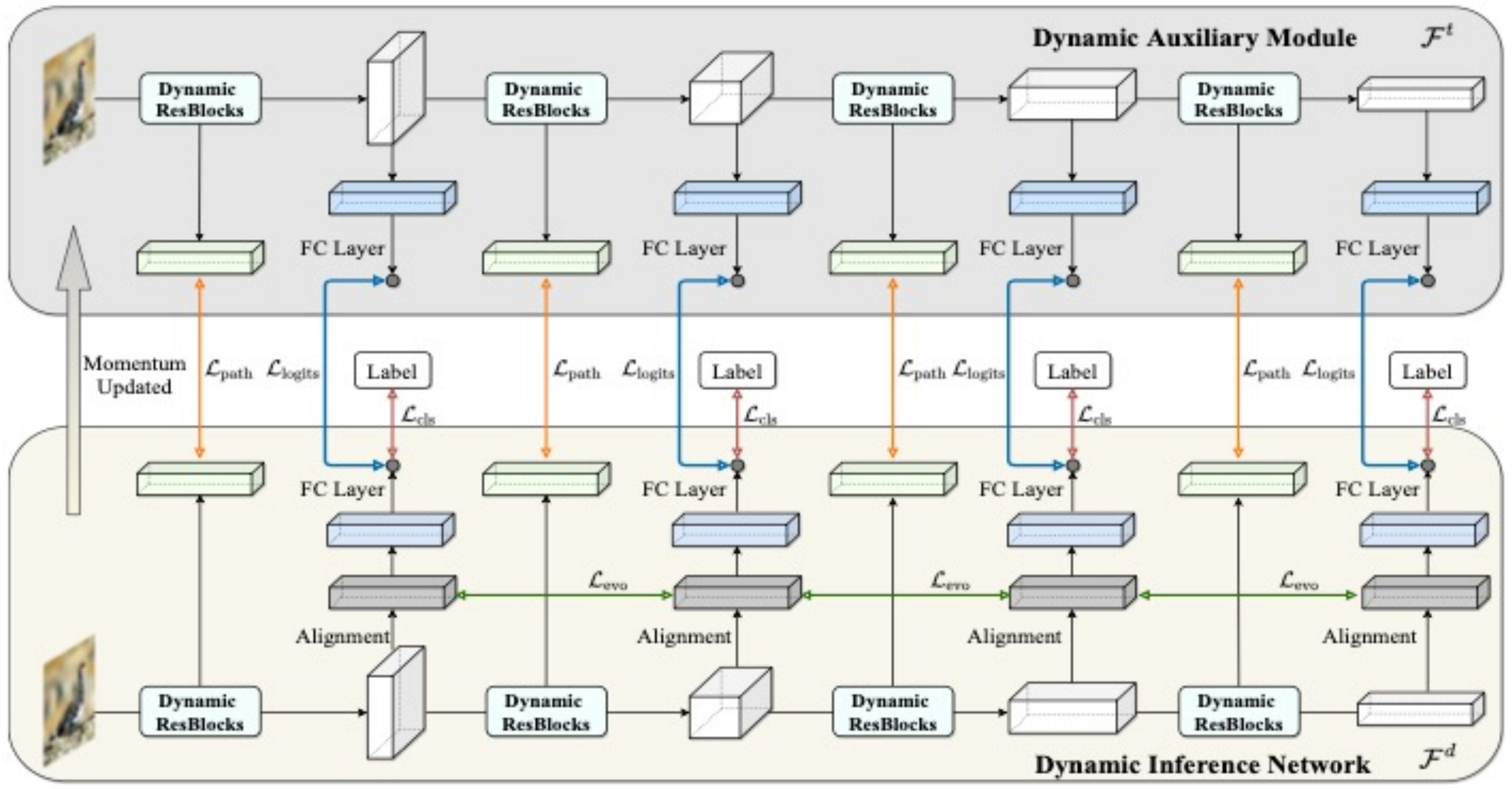
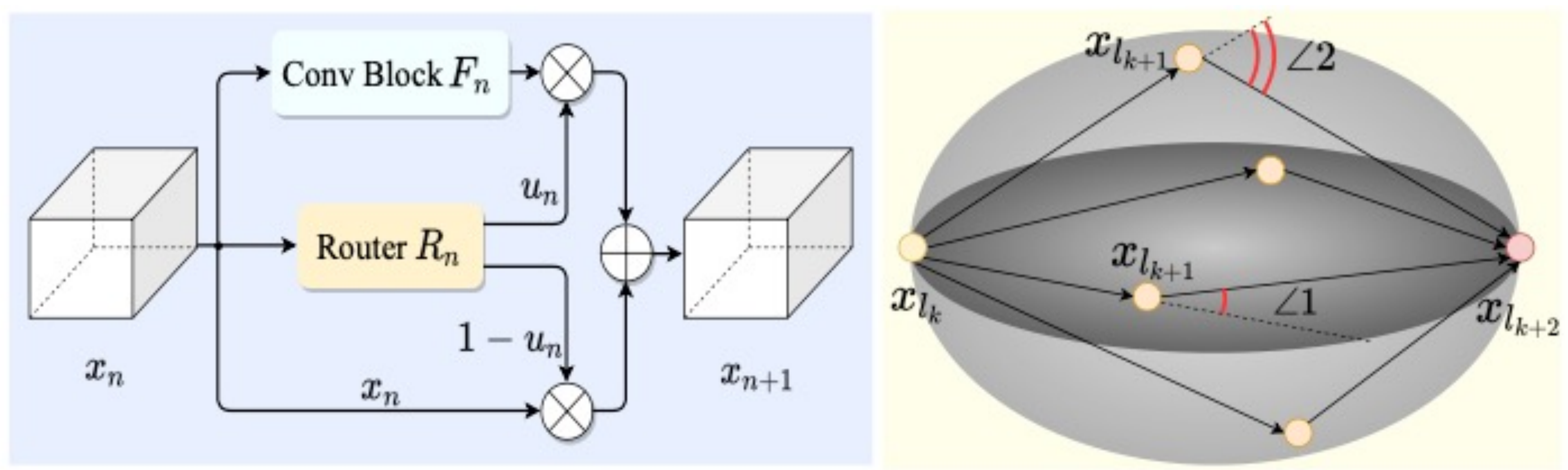


Illustration of the proposed Slowly Progressing Dynamic Inference Network. The proposed framework consists of a dynamic inference network and a dynamic auxiliary module. We attach several classifiers at different stages of the network.

In our method, we improve the optimization of the dynamic inference network from two aspects: utilizing the historical information and regularizing the direction of feature evolution. The whole framework of our method is shown in the above figure. Our proposed framework contains a dynamic inference network, a dynamic auxiliary module and several attached classifiers.

Feature Evolution Direction



(a) Dynamic Residual Block. (b) Feature Evolution.

(a) Illustration of the Dynamic Residual Block. A Dynamic Residual Block consist of a residual block and a router. (b) Illustration of regularizing on feature evolution. ∠1 is smaller than ∠2.

Experiments

Methods	Backbones	GMACCs	Acc. (%)
ResNet-32 [12]	—	0.14	92.40
ResNet-110 [12]	—	0.50	93.60
dynamic inference			
SkipNet [41]	ResNet-74	0.09	92.38
BlockDrop [38]	ResNet-110	0.17	93.60
ConvAIG [32]	ResNet-110	0.41	94.24
IamNN [19]	ResNet-101	1.10	94.60
CGap [4]	ResNet-110	0.19	93.43
CoDiNet [35]	ResNet-110	0.29	94.47
RDI-Net [34]	ResNet-110	0.38	95.10
early prediction			
ACT [5]	ResNet-110	0.38	93.50
SACT [5]	ResNet-110	0.31	93.40
DDI [36]	ResNet-74	0.14	93.88
DG-Net [27]	ResNet-101	3.20	93.99
DG-Net (light)	ResNet-101	2.22	91.99
SP-Net	ResNet-110	0.46	95.22
SP-Net (light)	ResNet-110	0.13	93.79

Comparison with state-of-the-arts on CIFAR-10

Methods	Backbones	GMACCs	Acc. (%)
ResNet-50 [12]	—	7.72	75.36
ResNet-101 [12]	—	15.26	76.45
dynamic inference			
ConvAIG [32]	ResNet-50	6.12	76.18
SkipNet [41]	ResNet-101	13.40	77.40
SkipNet (light)	ResNet-101	7.20	75.22
LCNet [39]	ResNet-50	5.78	74.10
BlockDrop [38]	ResNet-101	14.64	76.80
DG-Net [27]	ResNet-101	14.10	76.80
CoDiNet [35]	ResNet-50	6.20	76.63
RDI-Net [34]	ResNet-50	7.42	76.96
early prediction			
MSDN [15]	DenseNets	4.60	74.24
RA-Net [43]	DenseNets	4.80	75.10
IamNN [19]	ResNet-101	8.00	69.50
ACT [5]	ResNets	13.40	75.30
SACT [5]	ResNets	14.40	75.80
SP-Net	ResNet-50	7.24	77.21
SP-Net (light)	ResNet-50	5.62	76.41

Comparison with state-of-the-arts on ImageNet