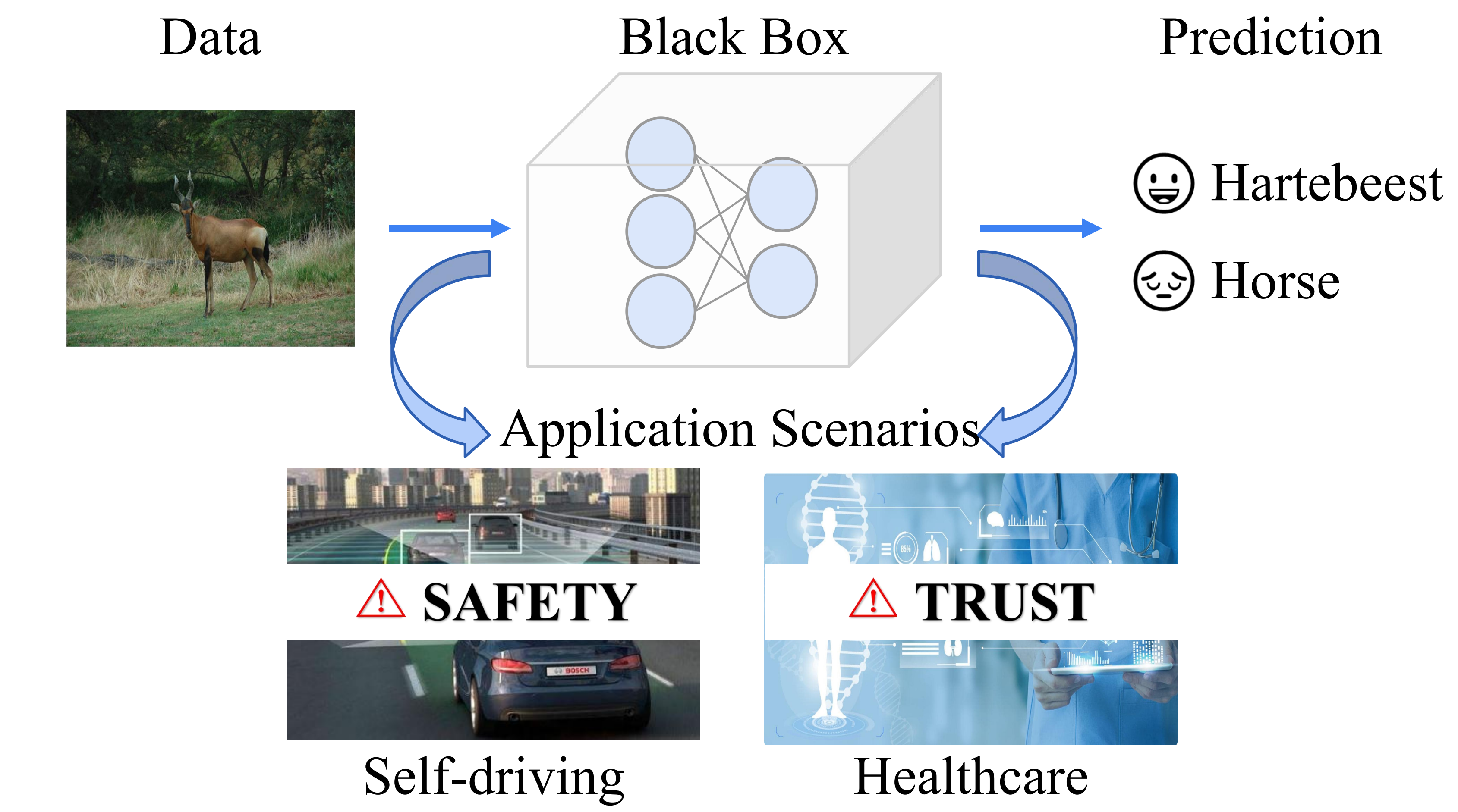
GitHub

Understanding the Dynamics of DNNs Using Graph Modularity

Yao Lu[1], Wen Yang[1], Yunzhe Zhang[1], Zuohui Chen[1], Jinyin Chen[1], Qi Xuan[1], Zhen Wang[2], Xiaoniu Yang[3]
1 Zhejiang University of Technology, 2 Northwestern Polytechnical University,
3 Science and Technology on Communication Information Security Control Laboratory

Paper

The Significance of Explainable Artificial Intelligence

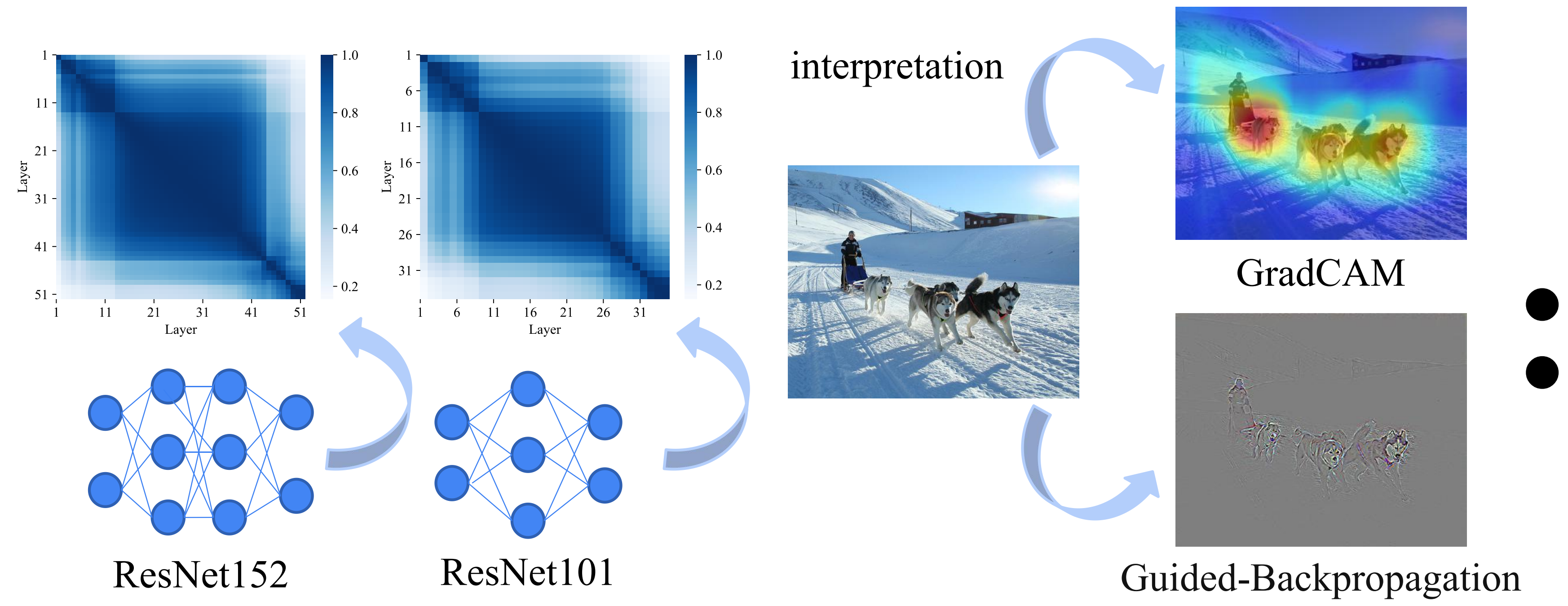


The black-box nature of DNNs hinders their applicability to high-stakes decision-making domains, such as healthcare, self-driving.

Previous Work on Characterizing Features

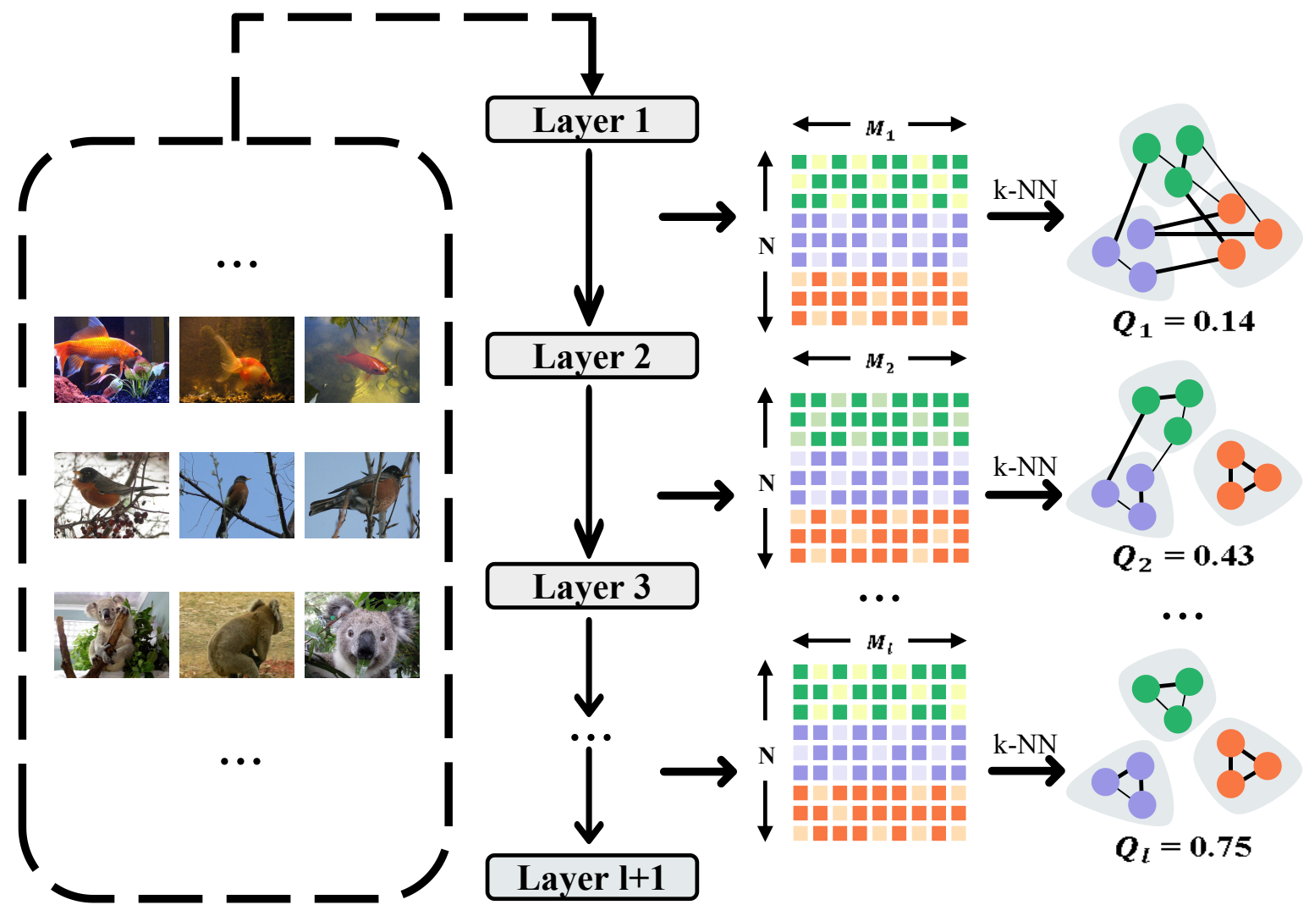
Measuring Representational Similarities

Interpreting Feature Semantics



Limitations: existing studies ignore the dynamics of DNNs or only understand the dynamics of DNNs through qualitative visualization.

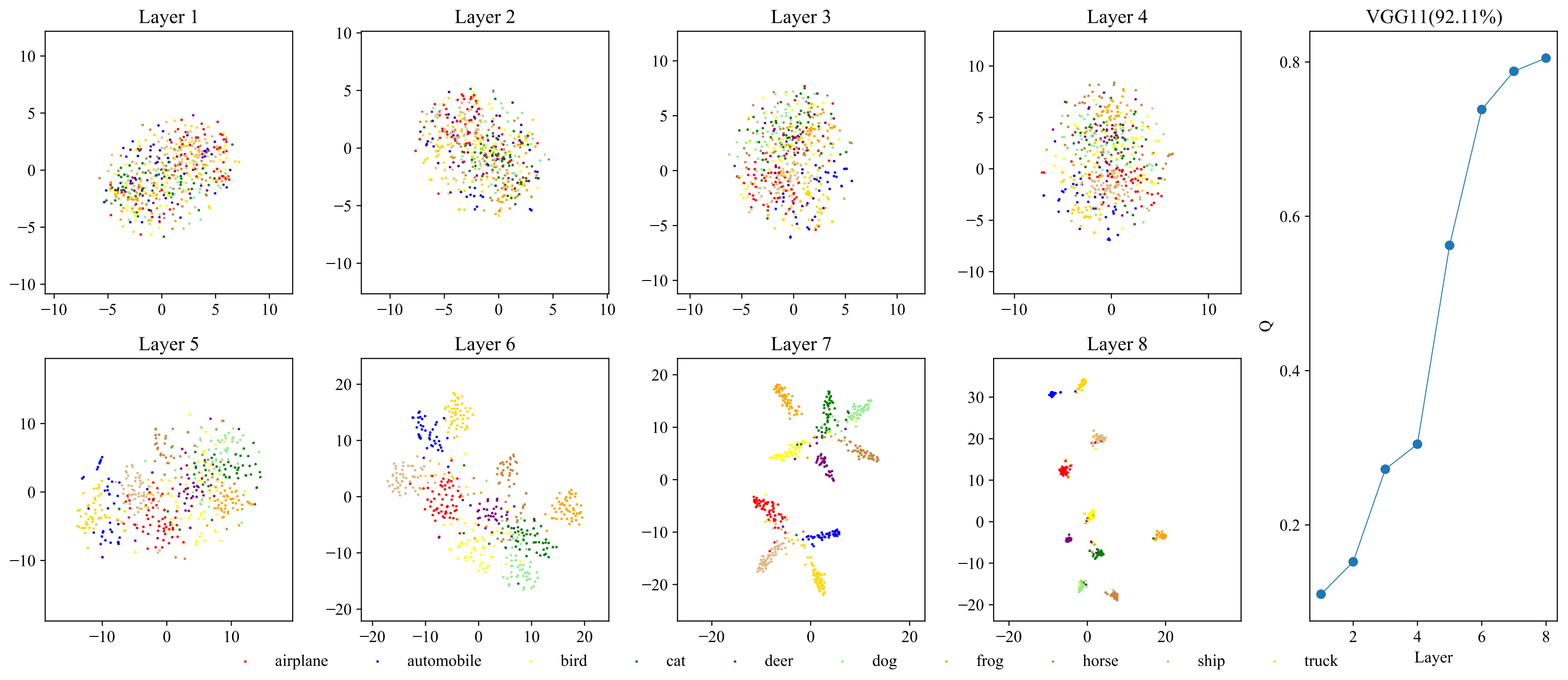
Dynamic Graph Construction



$$Q = \frac{1}{2W}\sum_{ij}\left(a_{ij} - \frac{s_i s_j}{2W}\right)\delta(c_i, c_j)$$

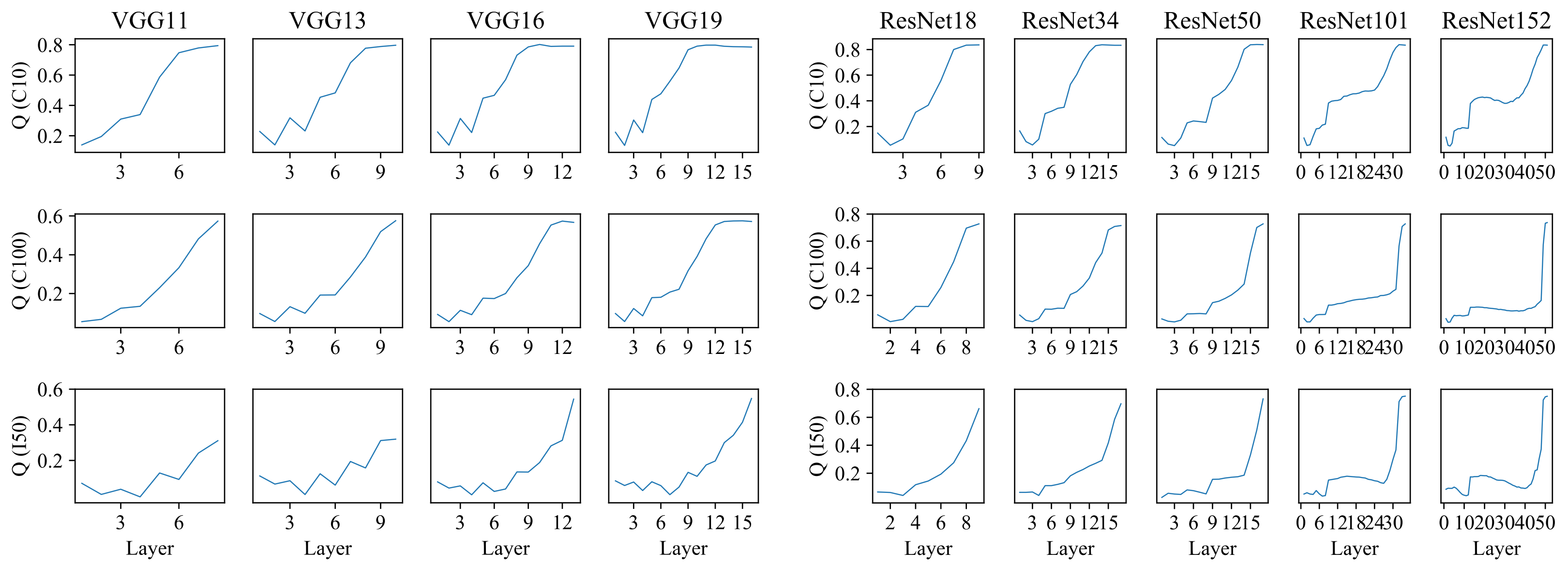
Pipeline for the dynamic graph construction and the corresponding calculation formula of modularity

Quantifying the Class Separation Process

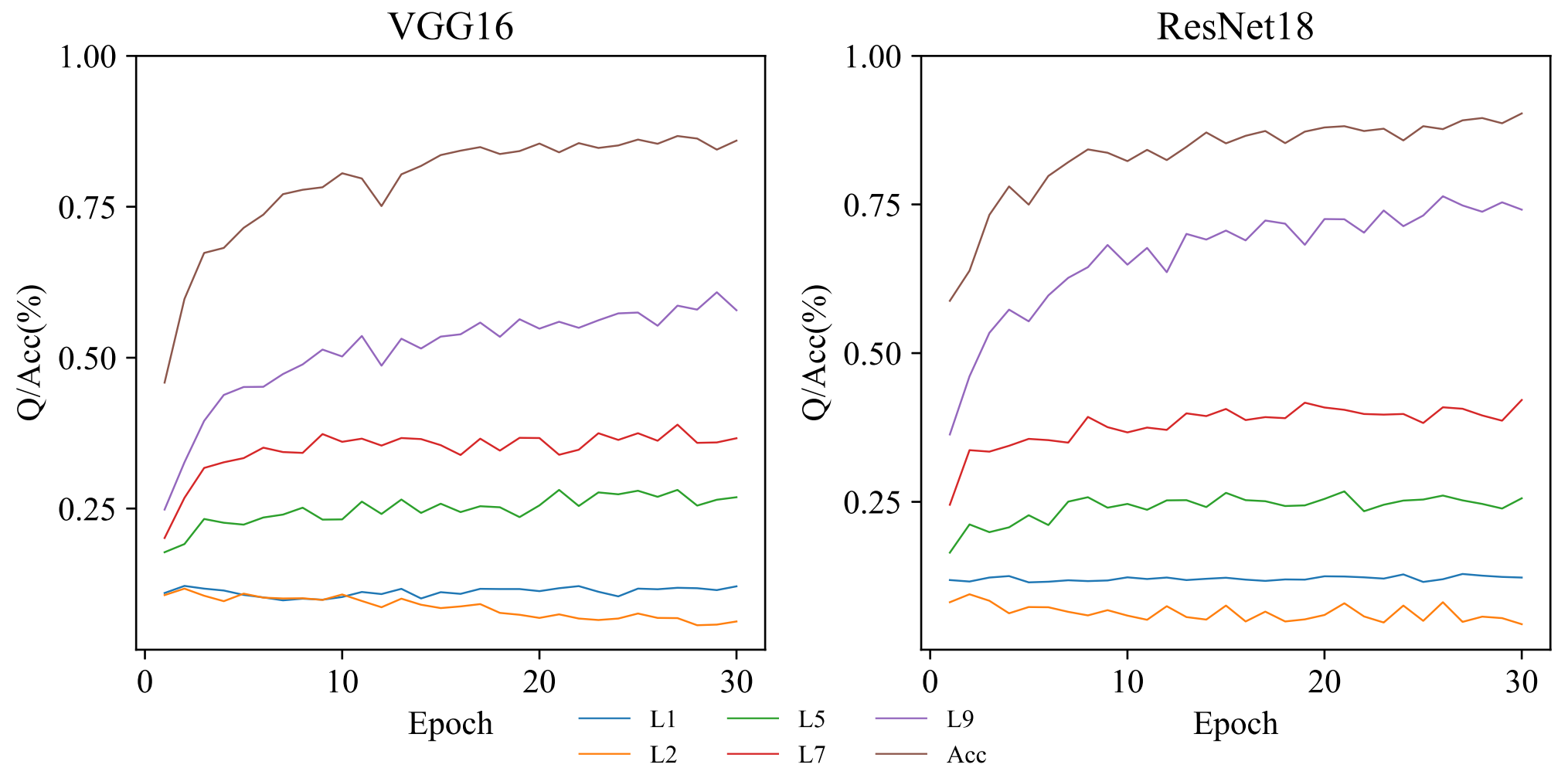


Modularity provides a quantifiable interpretation perspective for understanding the dynamics of DNNs

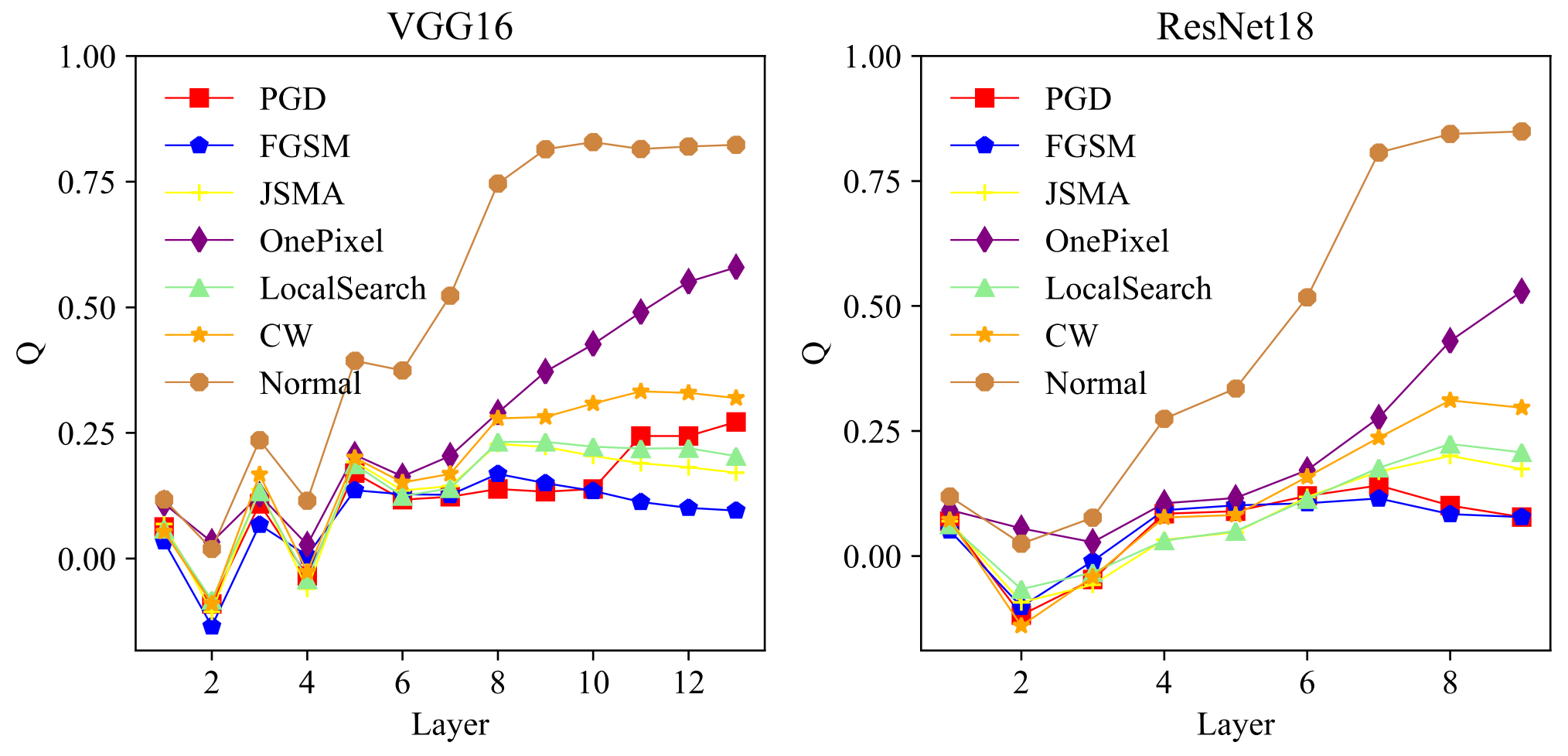
Modularity curves in different scenarios



- The modularity tends to increase as the layer goes deeper.
- Degradation and plateau are related to model relative complexity.

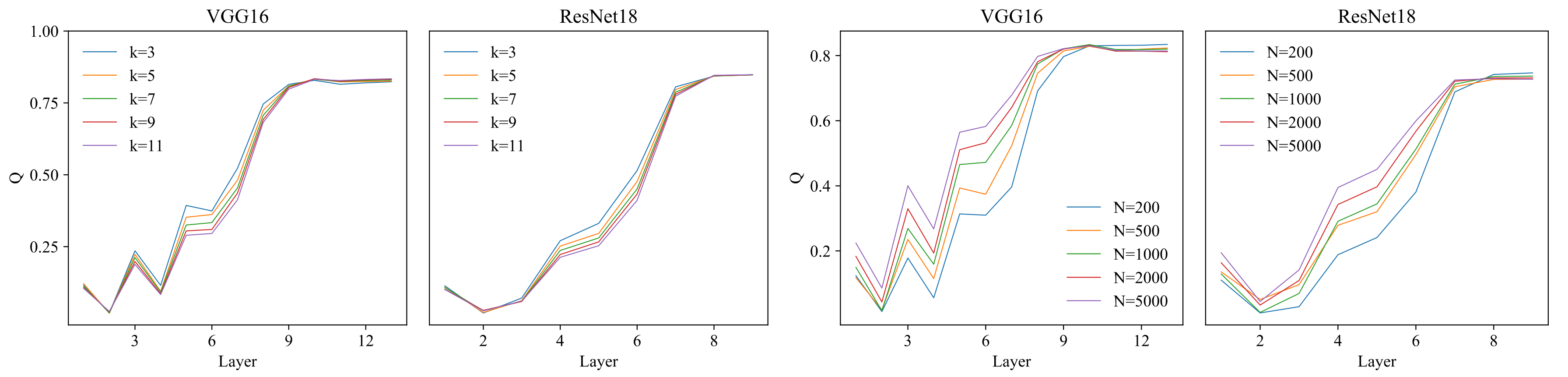


Shallow layers in DNNs extract general features while deep layers learn more specifically.



Adversarial attacks blur the distinctions among various categories.

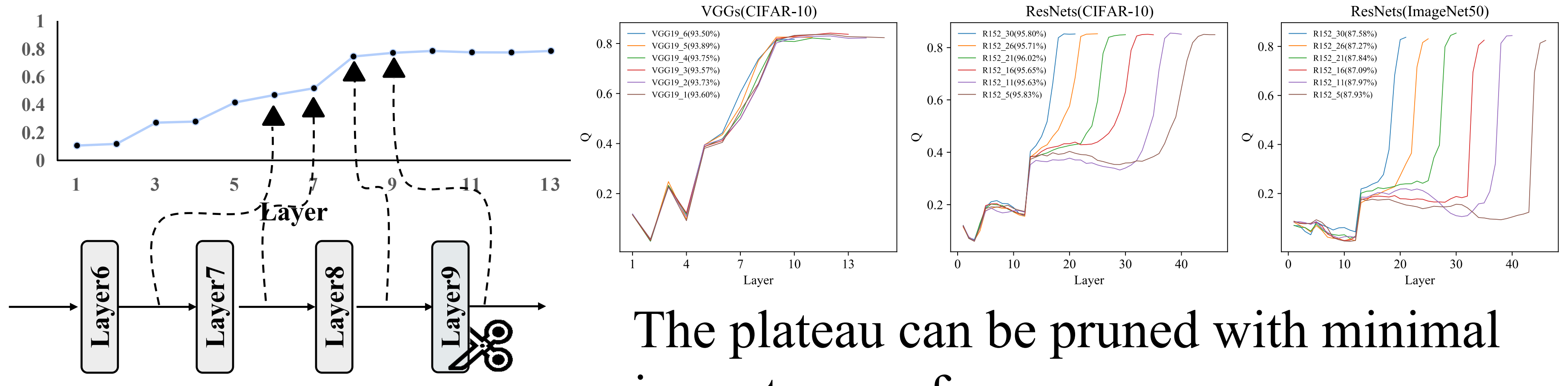
Ablation Study



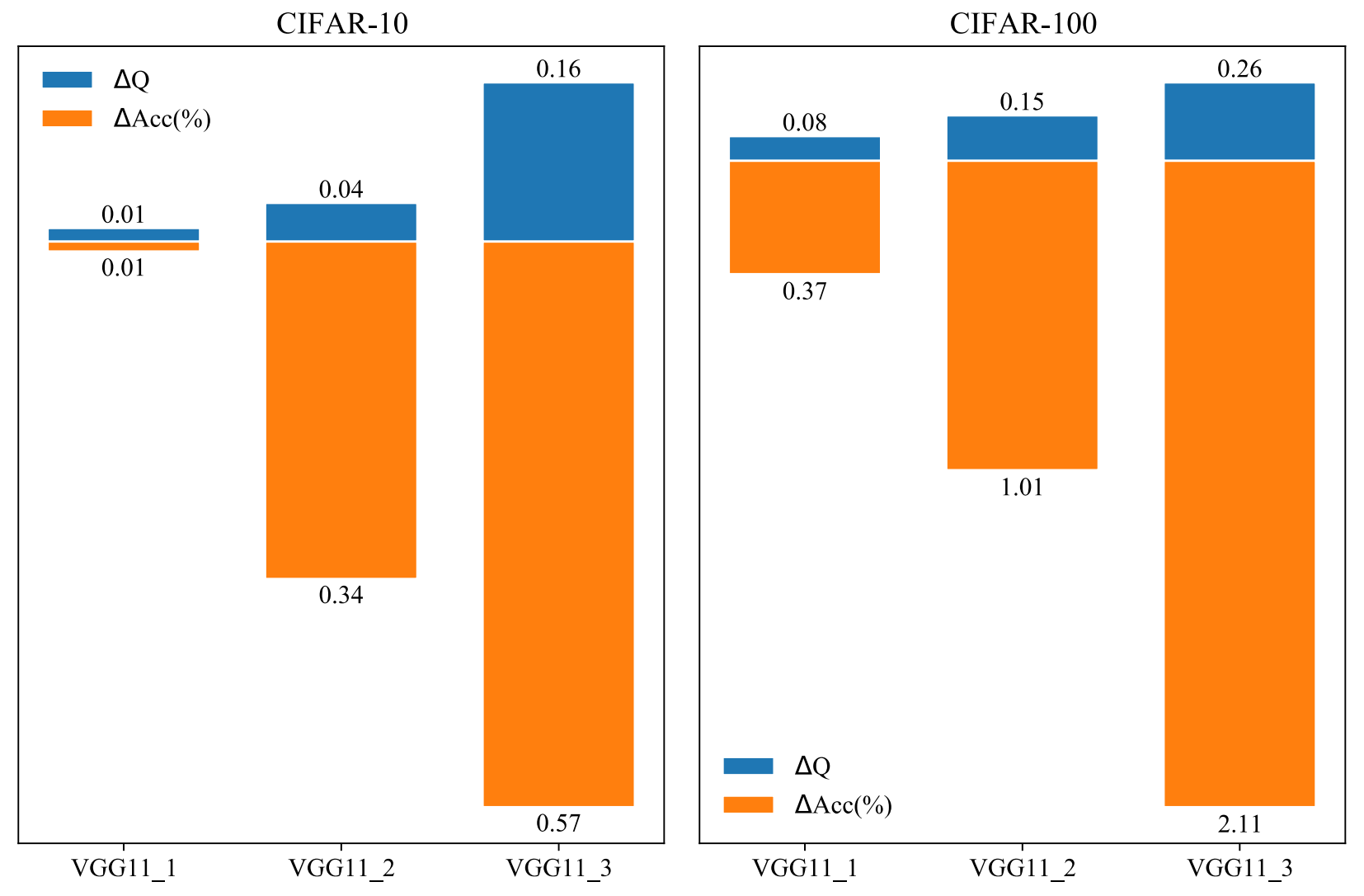
The modularity is reliable for hyperparameters.

Application Scenarios of Modularity

Guiding Layer Pruning



The plateau can be pruned with minimal impact on performance.

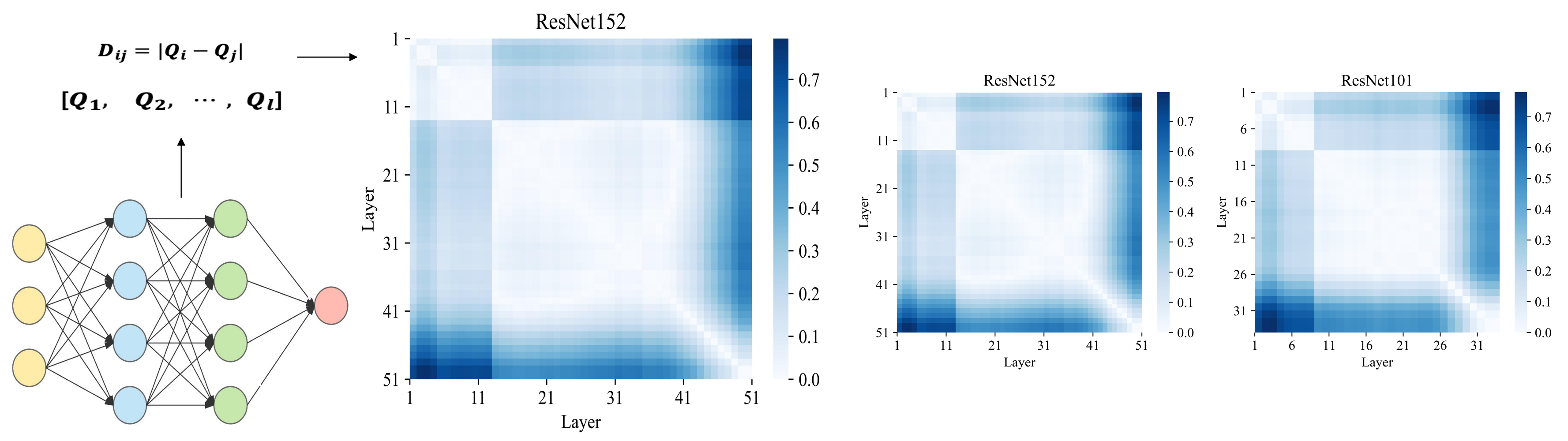


Pruning irredundant layers will result in a significant drop in performance.

Method	Top-1%	Params(PR)	FLOPs(PR)
ResNet56	93.27	0%	0%
Chen et al [5]	93.29	42.30%	34.80%
DBP-0.5 [61]	93.39	/	53.41%
Ours	93.38	43.00%	60.30%

Layer pruning with the guidance of modularity can achieve state-of-the-art performance.

Representing the Difference of Layers



Representations after residual connections are more different from that inside ResNet blocks than other post-residual ones.

INSTITUTE OF CYBERSPACE SECURITY

October 23-27, 2022, Tel Aviv

西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

CETC
中国电子科技集团公司第三十六研究所
THE 36TH RESEARCH INSTITUTE OF CHINA ELECTRONICS TECHNOLOGY GROUP CORPORATION