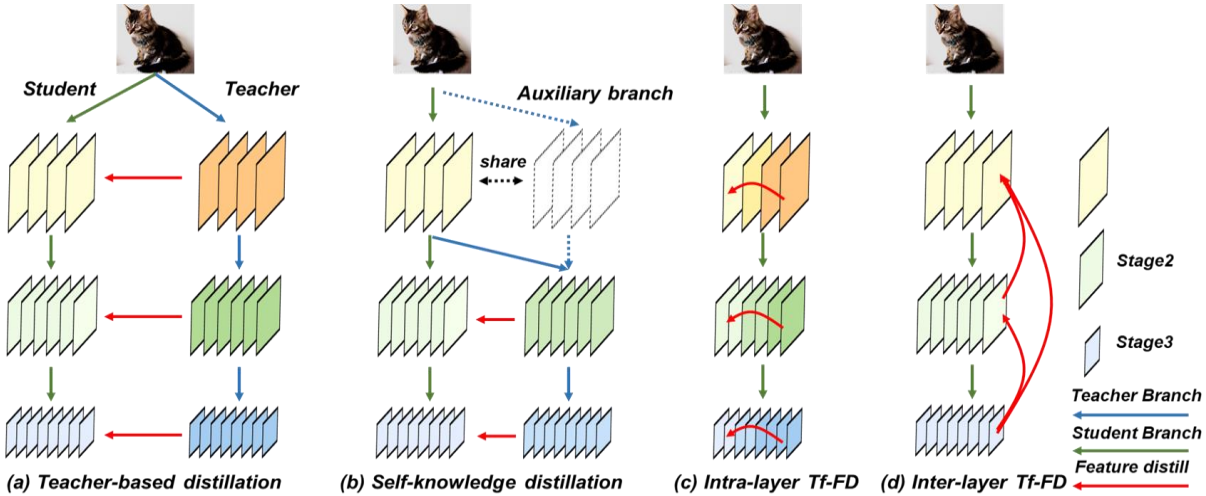
Self-Regulated Feature Learning via Teacher-free Feature Distillation

Lujun LI –lilujunai@gmail.com models and code are publicly available:https://lilujunai.github.io/Teacher-free-Distillation/

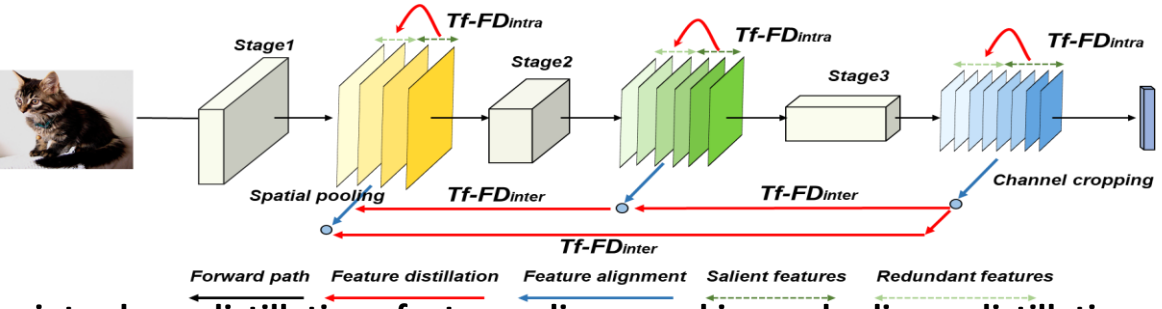
What role of the teacher model in knowledge distillation?

Cons (-): need to select & train teacher model, complex kd strategies



Intriguing observation: offline/online & strong/weak model & branch, channel, layer level self- features can all be used as teachers can be teachers

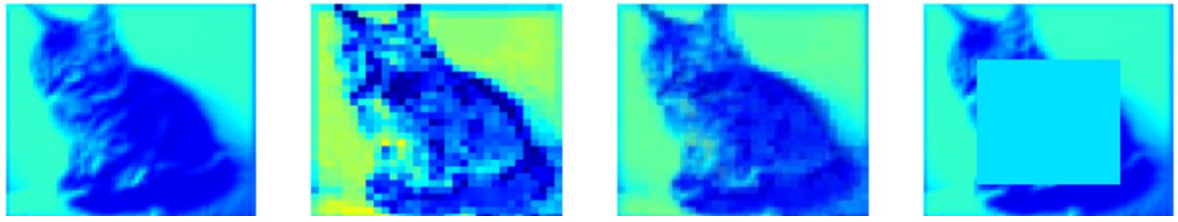
Teacher-free Feature Distillation



intra-layer distillation: feature saliency ranking and saliency distillation

inter-layer distillation: cross-layer feature alignment and distillation

Understanding from feature regularization



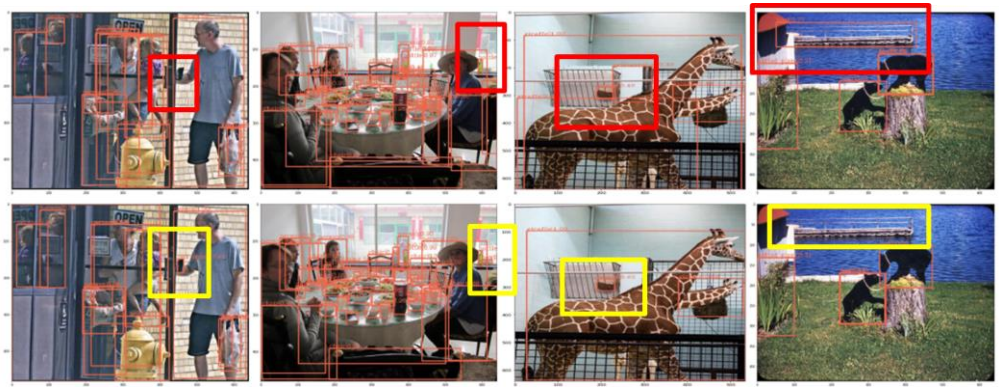
Experiments

Classification accuracy on Cifar-100 dataset

Model	R20	R32	W-16-2	R8-4	VGG8
Base acc	69.06	71.14	73.26	72.50	70.36
T*f*-FD	70.62	72.55	74.33	73.62	71.62
KD	70.67	73.08	74.92	73.33	72.98
KD+T*f*-FD	71.56	73.68	75.68	75.65	74.08

Results are also reported on ImageNet and MS-COCO dataset

T*f*-FD helps detectors to detect small objects and kill FP



Fast training speed & Good orthogonality with other distillations