ScratchDet: Training Single-Shot Object Detectors from Scratch

Rui Zhu*, Shifeng Zhang*, Xiaobo Wang, Longyin Wen, Hailin Shi†, Liefeng Bo, Tao Mei

JD AI

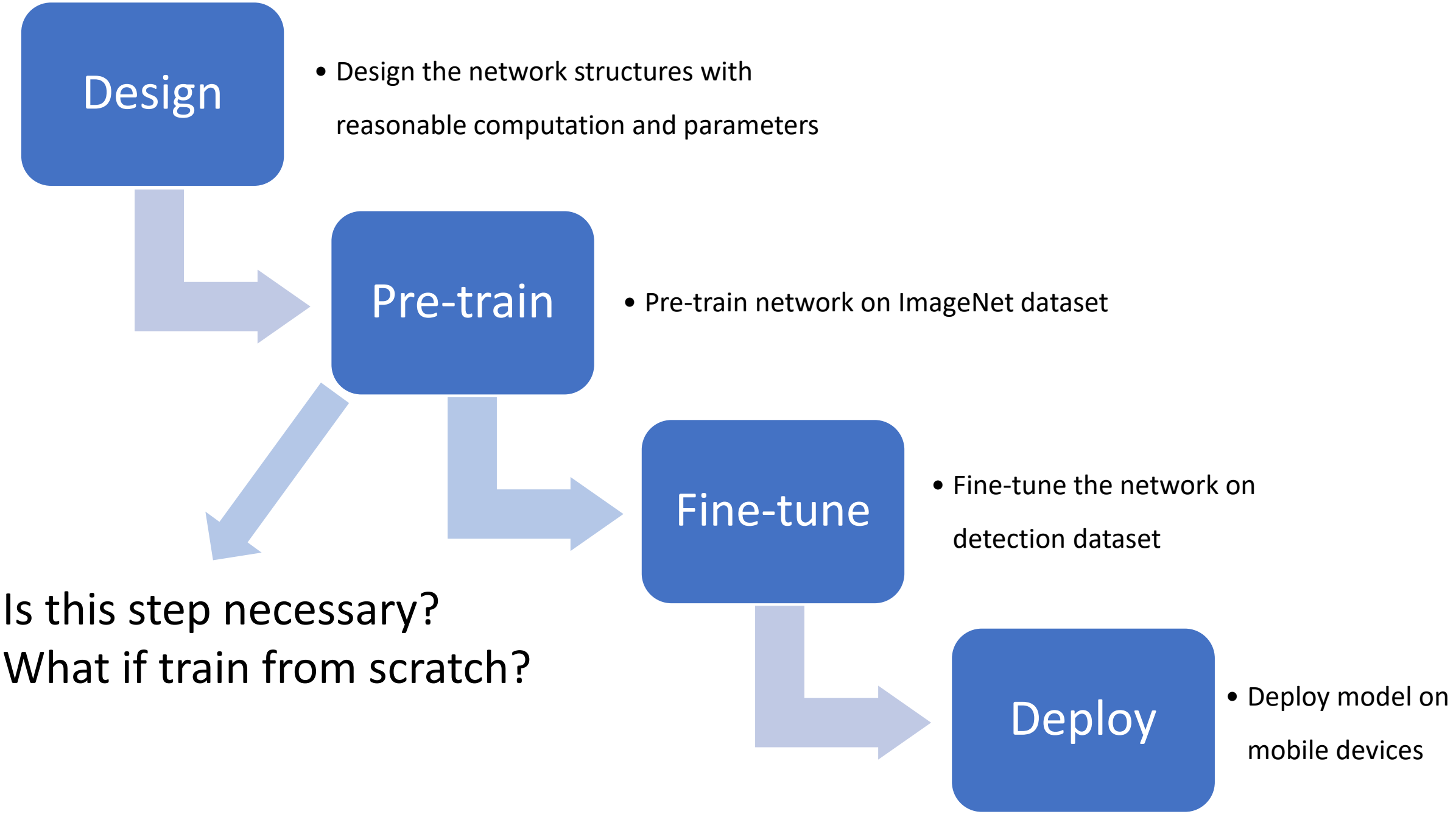
Institute of Automation, Chinese Academy of Science

JD Digits

Sun Yat-sen University

CVPR 2019, Long Beach, CA

How to design an efficient detection network for ARM/edge devices?


Design
• Design the network structures with reasonable computation and parameters
Pre-train
• Pre-train network on ImageNet dataset
Fine-tune
• Fine-tune the network on detection dataset
Deploy
• Deploy model on mobile devices
Is this step necessary?
What if train from scratch?

Fine-tuning or Training from scratch

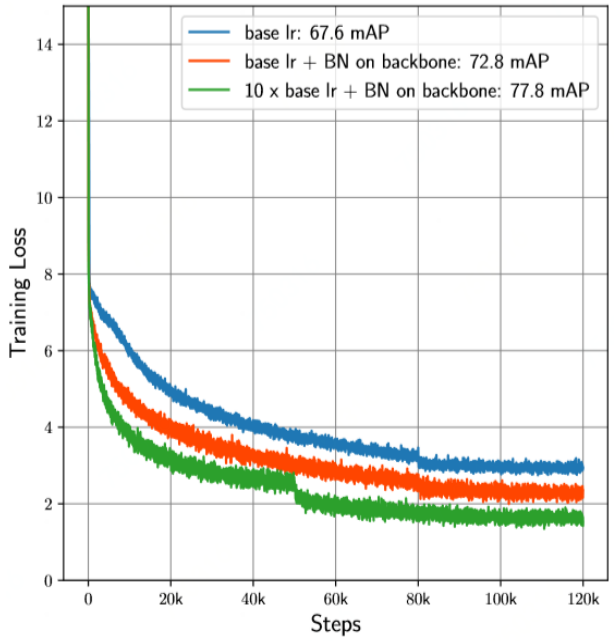
- Previous fine-tuning strategies:
  - *different degrees of sensitivity to translation*
  - *learning bias from classification to detection*
  - *inconvenient to change the architecture of networks (high computational cost on ImageNet)*
- Targets of training from scratch strategies:
  - *free the architecture limitations from classification*
  - *guarantee the training convergence*
  - *performance as good as fine-tune strategies*

Contributions:

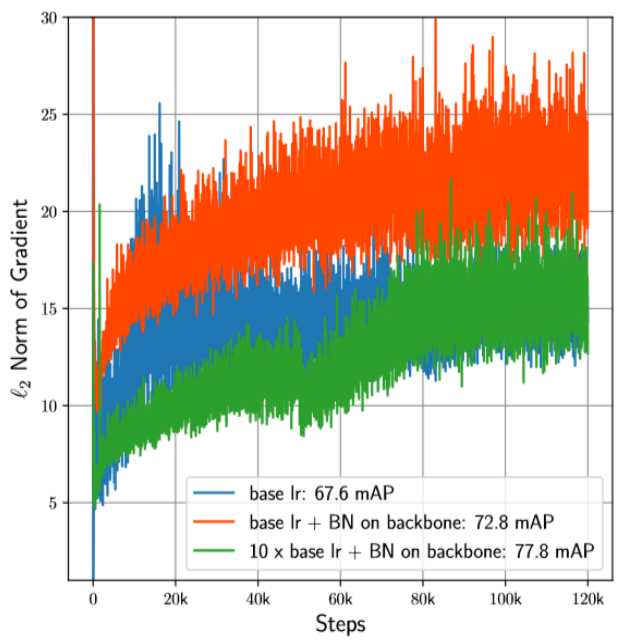
- Analyze the impact of BatchNormalization for train-from-scratch
- Structure design w/o pre-train for small object detection
- Extensive experiments on Benchmarks

The impact of BatchNorm for train-from-scratch:

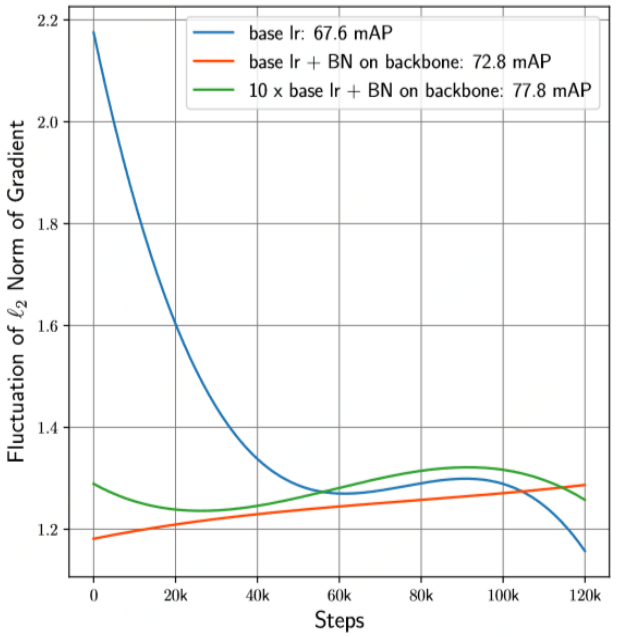
- smoother optimization landscape & more stable gradients
- enabling larger learning rate
- successfully train SSD from scratch
- free to modify network structure w/o restrictions of pre-train



(a) Loss Value



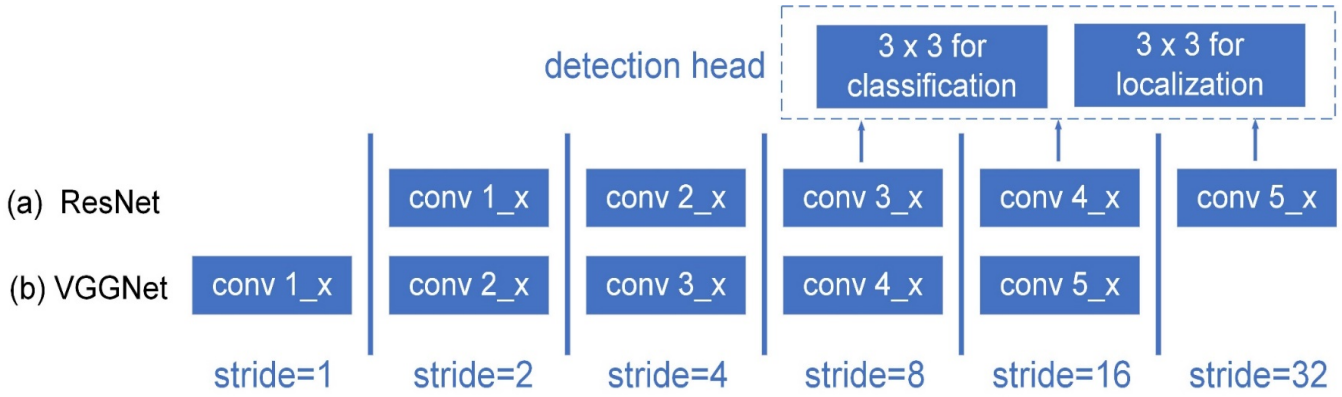
(b) L2 Norm of Gradient

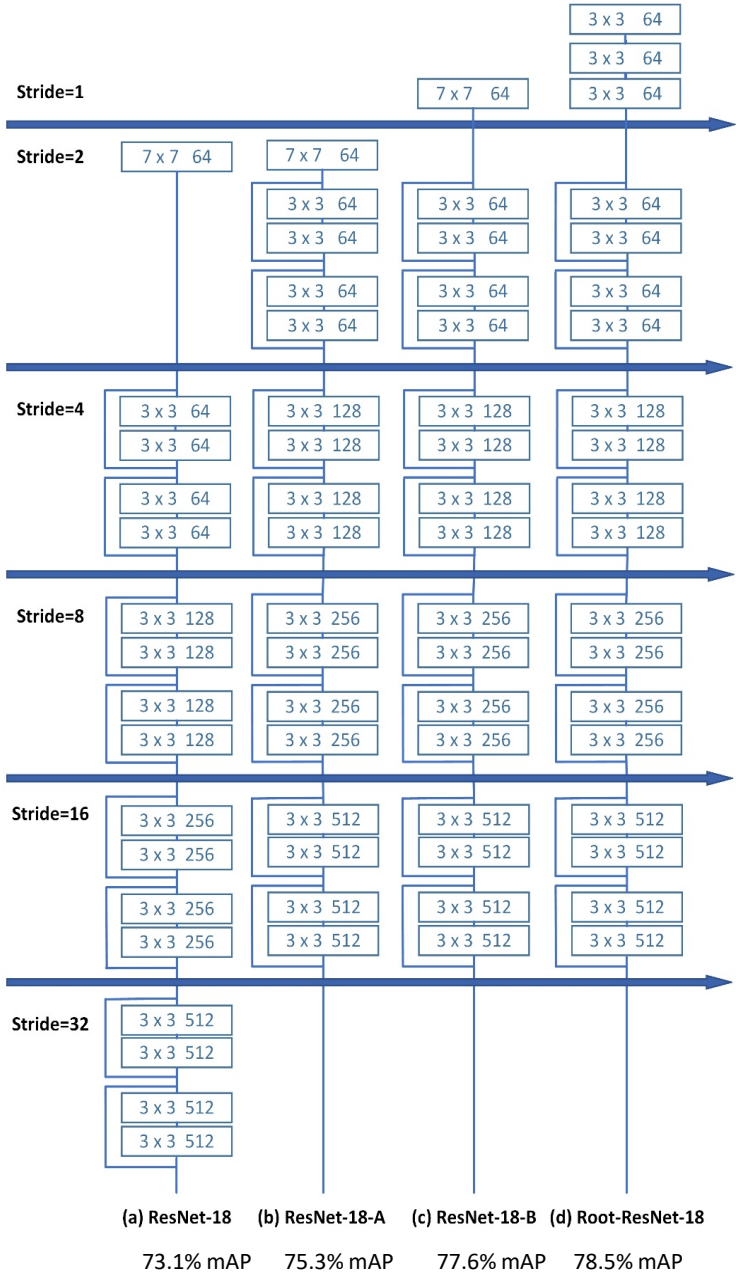


(c) Fluctuation of L2 Norm of Gradient

Structure redesign w/o pre-train

- Structure redesign for small object detection
- performance analysis of ResNet and VGGNet
- backbone network redesign for small objects detection
- good performance of Root-ResNet on SSD-300
- unnecessary to pre-train





Extensive experiments on Benchmarks

- Sufficient & fair exploratory experiments on VOC2007：

Model Analysis					
Component	**SSD-300-scratch**				
50x large learning rate?	✔				
10x large learning rate?		✔			
BN on backbone?	✔	✔	✔		
BN on detection head?	✔	✔	✔	✔	
mAP (%) on VOC07 test	**78.7**	**77.3**	**71.8**	**71.0**	**67.6**

Extensive experiments on Benchmarks

- Competitive performance on VOC2007, 2012 & MS COCO:
- Comparison of training time:

  Scratch (84.6h) vs. Fine-tune (29.7h)

  Attractive time-saving considering several weeks pre-train

Results on Benchmarks of ScratchDet-300							
Method	**VOC07 test**		**VOC12 test**		**COCO test-dev**		
	0712	**0712+ COCO**	**0712**	**0712+ COCO**	**Trainval35k**		
	0.5	**0.5**	**0.5**	**0.5**	**0.5:0.95**	**0.5**	**0.75**
Single Test	80.4	84.0	78.5	82.1	32.7	52.0	34.9
Multi Test	84.1	86.3	83.6	86.3	39.1	59.2	42.6

Conclusions:

- Study the effects of BatchNorm in the backbone and detection head subnetworks, and successfully train detectors from scratch.
- Be able to explore various architectures for detector designing.
- Propose a new Root-ResNet backbone network to further improve the detection accuracy, especially for small objects.
- Codes and models: *https: //github.com/KimSoybean/ScratchDet*

Thank You ！