

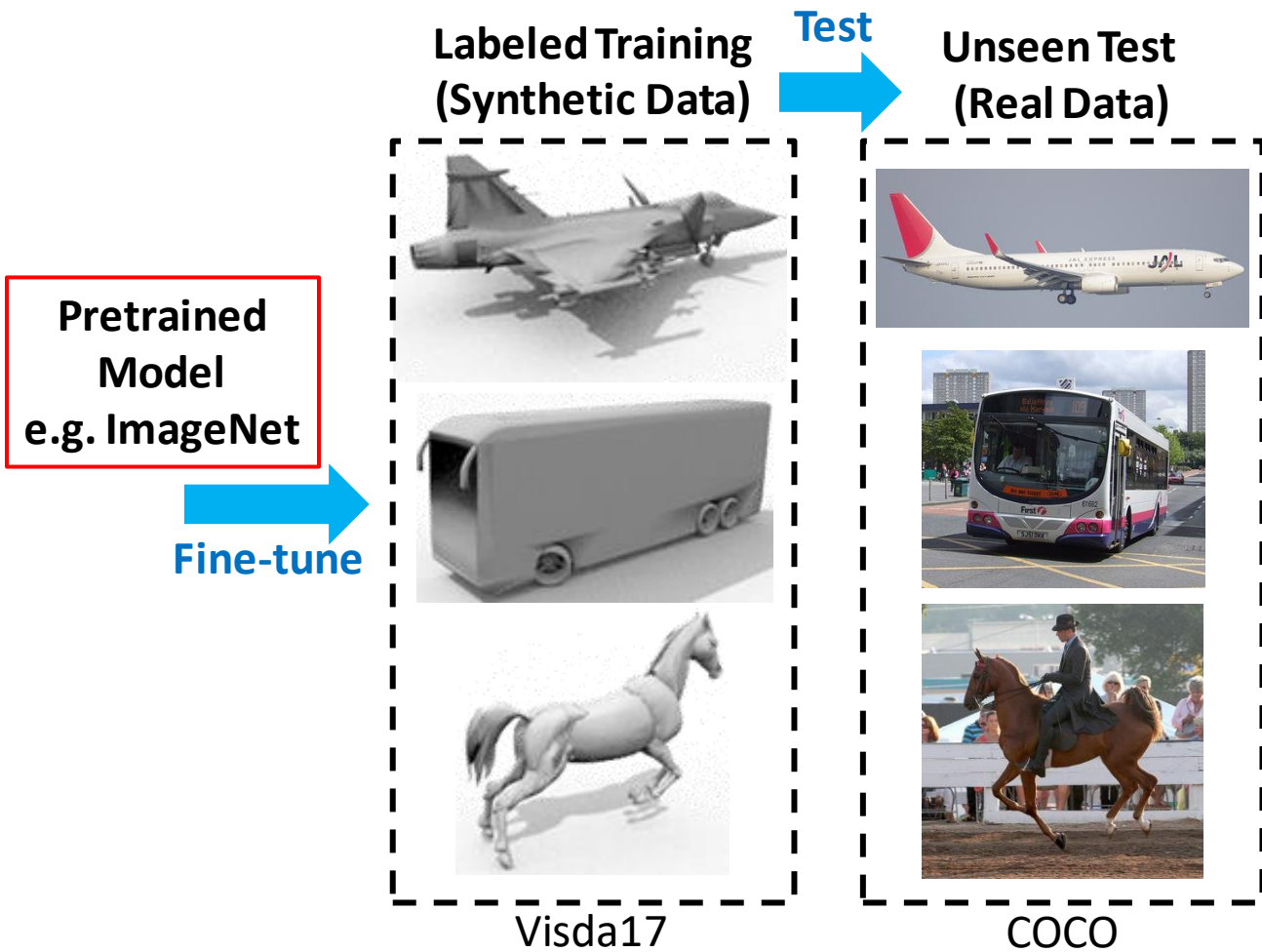
Automated Synthetic-to-Real Generalization

Wuyang Chen^{1*}, Zhiding Yu²,
Zhangyang Wang¹, Anima Anandkumar^{2 3}

¹ Texas A&M University ² NVIDIA ³ Caltech

* Work done during internship at NVIDIA

Syn-to-Real Generalization: Problem & Challenge

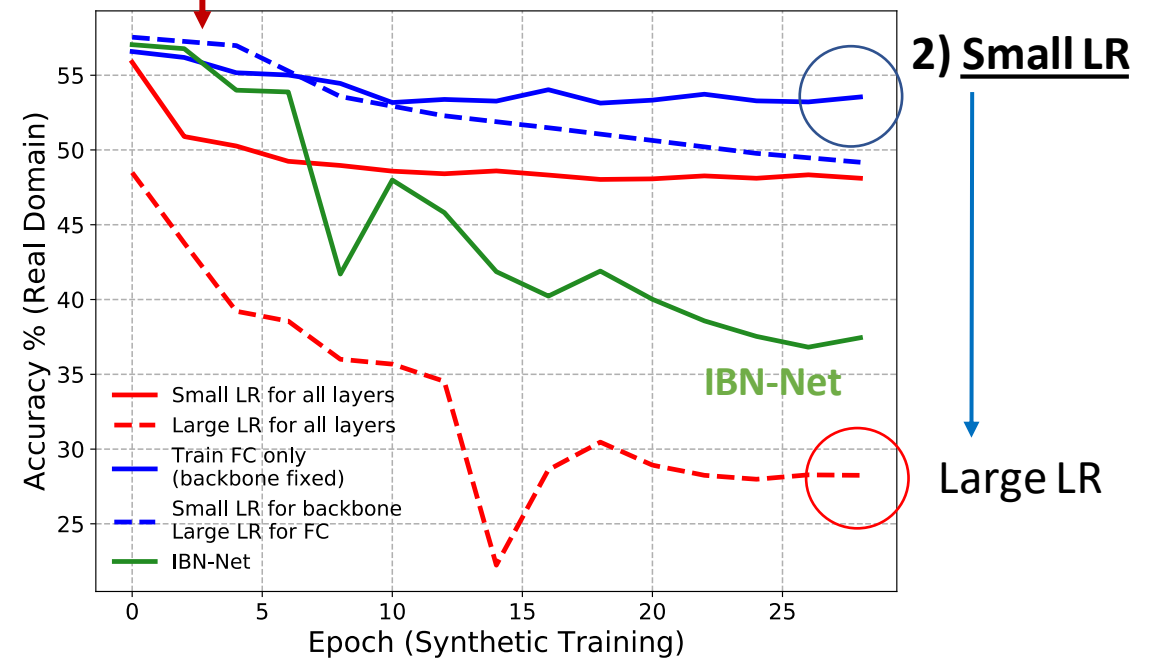


Challenge:

Large domain gap b/w synthetic and real data

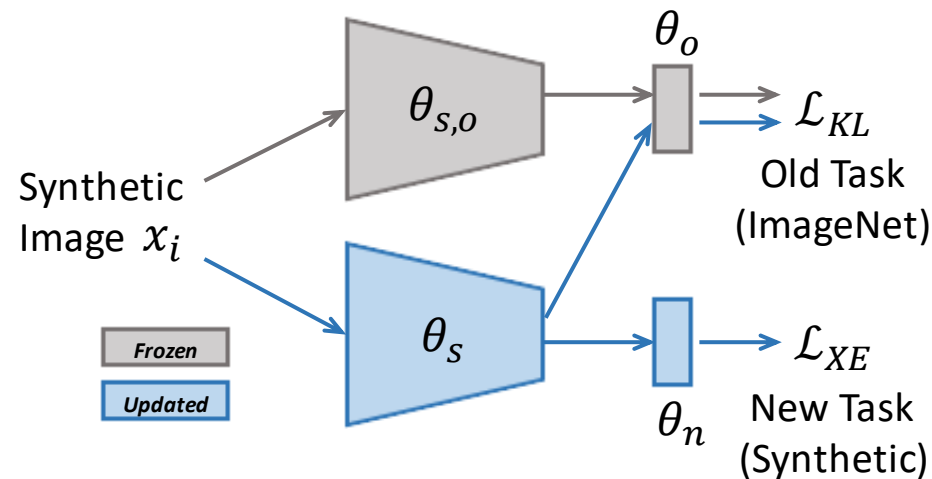
Prior solutions:

1) Early stopping



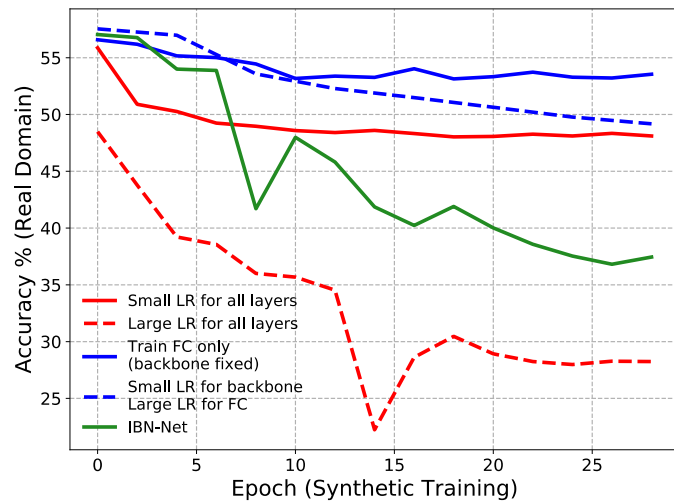
Syn-to-Real Generalization with Proxy Guidance

- Why people do **early stopping**?
 - ➔ Do not train too far from initialization (ImageNet pretrained weight).
- We minimize KL divergence \mathcal{L}_{KL} b/w new model and initialization.
 - ➔ ImageNet pretrained weight as proxy guidance in syn2real training.

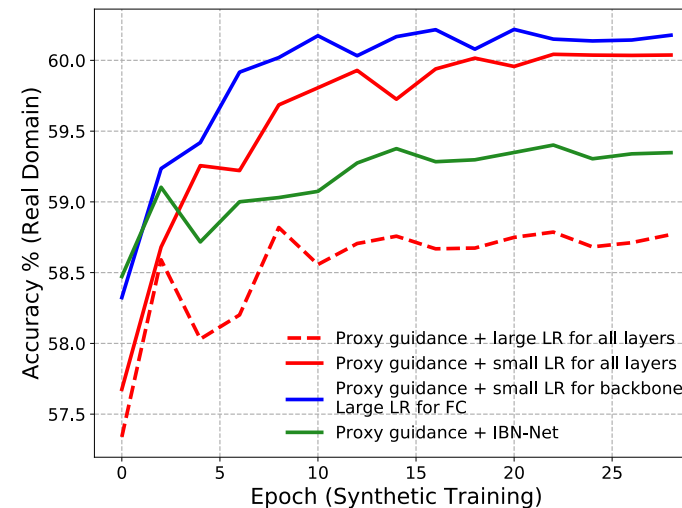


Syn-to-Real Generalization with Proxy Guidance

- Why people do **early stopping**?
 - ➔ Do not train too far from initialization (ImageNet pretrained weight).
- We minimize KL divergence \mathcal{L}_{KL} b/w new model and initialization.
 - ➔ ImageNet pretrained weight as proxy guidance in syn2real training.



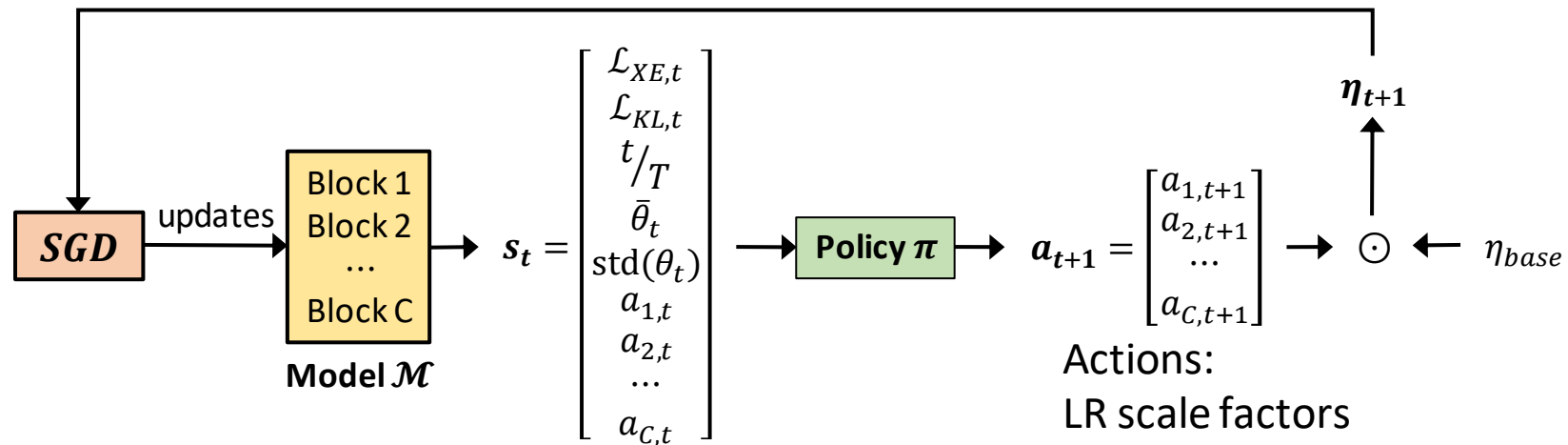
Proxy Guidance



Visda17 → COCO

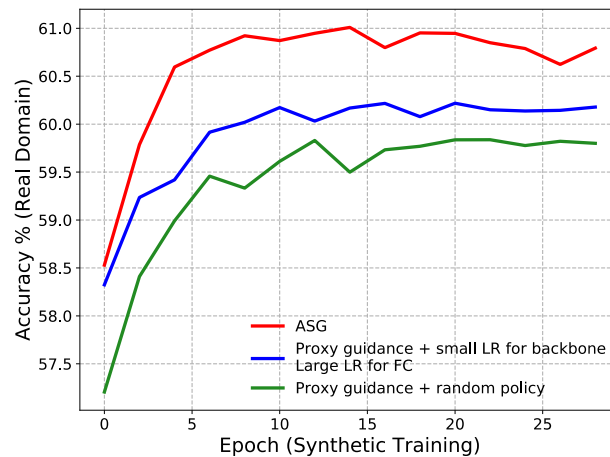
Automated Synthetic-to-Real Generalization (ASG)

- Why people use small learning rate?
 - ➔ Carefully fine-tune to avoid being far from initialization (ImageNet pretrained weight).
- But how small for which layer?
 - ➔ **L2O** (learning-to-optimize): automatic control of layer-wise learning rate.
 - Train L2O policy π with REINFORCE to produce learning rate actions.

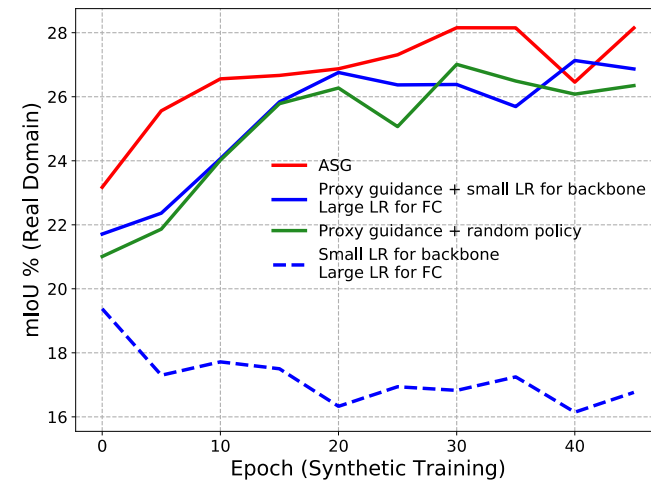


Automated Synthetic-to-Real Generalization (ASG)

- Why people use small learning rate?
 - ➔ Carefully fine-tune to avoid being far from initialization (ImageNet pretrained weight).
- But how small for which layer?
 - ➔ **L2O** (learning-to-optimize): automatic policy control of layer-wise learning rate.



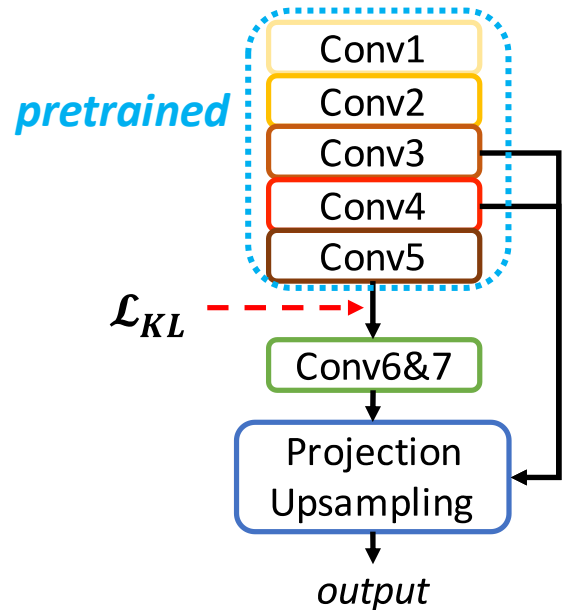
Classification: Visda17 ➔ COCO



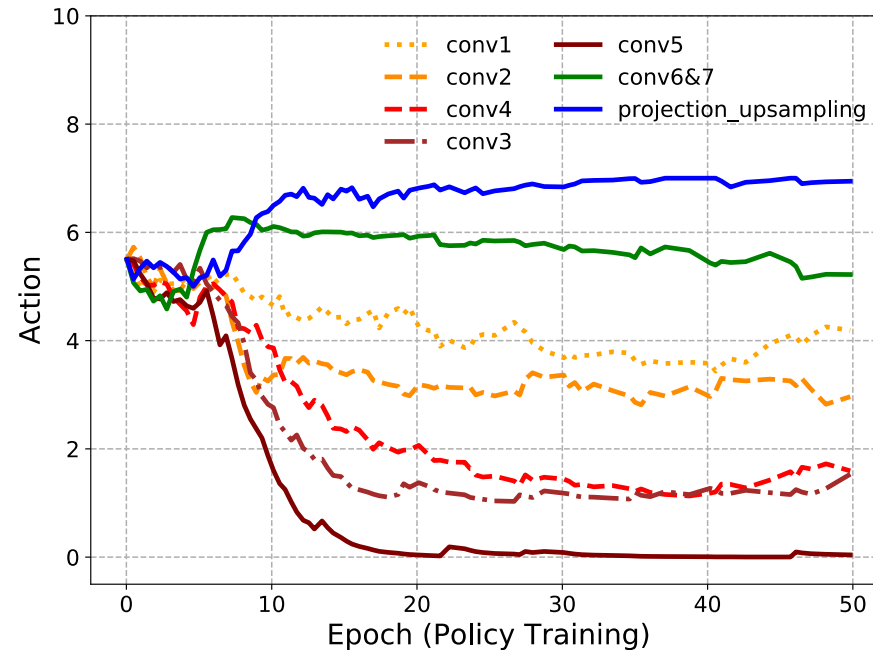
Segmentation: GTA5 ➔ Cityscapes

Action Behavior of RL-L2O Policy

- Backbone (ImageNet pretrained): closer to \mathcal{L}_{KL} \rightarrow smaller LR.
- Projection head: large LR.



Vgg16-FCN8s



projection head
 conv6&7
 conv1
 conv2
 conv3,4
 conv5

Segmentation: GTA5 \rightarrow Cityscapes

Why ASG Works? Retaining ImageNet Information

- ImageNet is a large-scale real-world dataset, it provides rich information about real domain.

#	Model	Visda-17	ImageNet	
1.	Large LR for all layers	28.2	0.8	←
2.	+ our Proxy Guidance	58.7 (+30.5)	76.2 (+75.4)	←
3.	Small LR for backbone and large LR for FC	49.3	33.1	←
4.	+ our Proxy Guidance	60.2 (+10.9)	76.5 (+43.4)	←
5.	Oracle on ImageNet ²	53.3 (+4.0)	77.4	←
6.	ROAD (Chen et al., 2018)	57.1 (+7.8)	77.4	←
7.	Vanilla L2 distance	56.4 (+7.1)	49.1	←
8.	SI (Zenke et al., 2017)	57.6 (+8.3)	53.9	←
9.	ASG (ours)	61.5	76.7	←

ASG Benefits Domain Adaptation

- ASG as initialization for domain adaptation methods.

Method	Tgt Img	Accuracy
Source (Saito et al., 2017)	✗	52.4
MCD (Saito et al., 2018)	✓	71.9
ADR (Saito et al., 2017)	✓	74.8
SimNet-Res152 (Pinheiro, 2018)	✓	72.9
GTA-Res152 (Sankaranarayanan et al., 2018)	✓	77.1
Source-Res101 (Zou et al., 2019)	✗	51.6
CBST (Zou et al., 2018)	✓	76.4 (0.9)
MRKLD (Zou et al., 2019)	✓	77.9 (0.5)
MRKLD + LRENT (Zou et al., 2019)	✓	78.1 (0.2)
ASG (ours)	✗	61.5
ASG + CBST	✓	82.5 (0.7)
ASG + MRKLD	✓	84.6 (0.4) ←
ASG + MRKLD + LRENT	✓	84.5 (0.4)

ASG Improves Model Attention (GradCAM)

Input Image



Baseline



Skateboard ✗

ASG



Airplane ✓



Train ✗



Bus ✓



Motorcycle ✗



Horse ✓

Thank you!

