Supplementary Materials: MutualNet: Adaptive ConvNet via Mutual Learning from Network Width and Resolution

Taojiannan Yang[1], Sijie Zhu[1], Chen Chen[1], Shen Yan[2], Mi Zhang[2], and Andrew Willis[1]

[1] University of North Carolina at Charlotte
{tyang30,szhu3,chen.chen,arwillis}@uncc.edu
[2] Michigan State University
{yanshen6,mizhang}@msu.edu

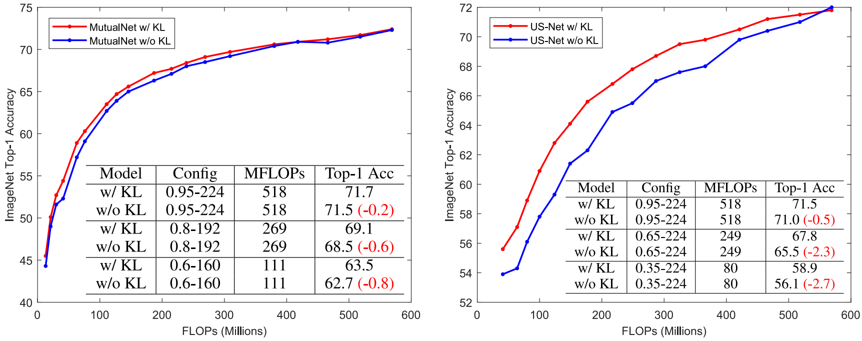
Abstract. This supplementary material includes the following items.
1. Contribution of KL divergence.
2. More object detection and instance segmentation examples.

1 Contribution of KL Divergence

In the experiments, we follow US-Net to train sub-networks using KL divergence (i.e., inplace distillation) to have a fair comparison. In US-Net, the authors claim that inplace distillation is an essential component for training US-Net. In this section, we study the contribution of inplace distillation to the overall performance of MutualNet. The training setting is the same as training on ImageNet, except that sub-networks are trained with the ground truth labels rather than the soft labels from the full-network. As shown in Fig. 1, training with KL divergence (i.e., inplace distillation) only achieves marginal improvement on MutualNet, while it is significant on US-Net. Without inplace distillation, the performance of US-Net drops around 2%. The stable performance of our MutualNet (w/ or w/o inplace distillation) is a clear advantage, which attributes to our proposed mutual learning scheme. As stated in the paper, each sub-network can leverage the knowledge learned by other sub-networks from different resolutions. *The mixed-resolution gradients in MutualNet can effectively transfer knowledge across every sub-network without resorting to knowledge distillation.* While US-Net has to rely on inplace distillation to transfer the knowledge from the full-network to sub-networks. This attribute of our (width and resolution) mutual learning scheme makes it possible to further simplify MutualNet training by removing inplace distillation. In this case, our MutualNet is easily applicable to the tasks where knowledge distillation is hard to apply (e.g., detection).

2 More Detection and Segmentation Examples

In this section, we provide more visual examples of object detection and instance segmentation in Fig. 2. We can clearly see that MutualNet can robustly detect



Model	Config	MFLOPs	Top-1 Acc
w/ KL	0.95-224	518	71.7
w/o KL	0.95-224	518	71.5 (-0.2)
w/ KL	0.8-192	269	69.1
w/o KL	0.8-192	269	68.5 (-0.6)
w/ KL	0.6-160	111	63.5
w/o KL	0.6-160	111	62.7 (-0.8)

(a) MutualNet

Model	Config	MFLOPs	Top-1 Acc
w/ KL	0.95-224	518	71.5
w/o KL	0.95-224	518	71.0 (-0.5)
w/ KL	0.65-224	249	67.8
w/o KL	0.65-224	249	65.5 (-2.3)
w/ KL	0.35-224	80	58.9
w/o KL	0.35-224	80	56.1 (-2.7)

(b) US-Net

Fig. 1: Contribution of KL divergence on the overall performance.

small scale and large scale objects while US-Net fails in some cases. This further demonstrates that MutualNet is able to capture robust multi-scale representations from network width and resolution.

Fig. 2: Object detection and instance segmentation examples.