To Dear Prof. Lin and Dear Prof. Loy July 2, 2017
Information Engineering, CUHK

Dear Professors,

I admire MMLAB a lot and want to devote diligently. May I state my tentative research interest?

Attention Mechanism:

From signal processing perspective, Nyquist–Shannon Sampling Theorem (Fixed sample frequency) has been a traditional rule. Guided by attention, we utilize the sparsity of real-world signals and extract meaningful semantic messages.

TSN [1] sparsely samples video snippets from segments and then weight the importance according to its content, which is equivalent to non-uniform sampling if following a re-sample step. But random sparse sample may miss some important information. Can we sample weight and re-sample iteratively or predict locations of important snippets while make sure the network can be trained end-to-end?

Scale-friendly Detection [2] use multi-scale classifiers and shared features. We may ask why simply fusing feature map still cannot prevent performance drop? Before reading [2], I am thinking a way to choose the most important feature map adaptively from redundant different scales and use a shared classifier. [2] proposes a excellent idea to make sure network differentiable! Meanwhile, we can refer to Deformable Conv [5] using data-driven attention to guide Pool and Conv so that receptive field change adaptively while the attention (offset field) is differentiable with regard to parameters.

DNN Design:

PolyNet [4] explore the structure diversity in polynomial form and there may be other structure diversity to explore, such as where to branch and where to merge. In spite of great expressive power, DNN still needs some hand-crafted and explicit design for specific task. This research topic is difficult but valuable. Google Brain [6] use Reinforce Learning to train a Policy Network(A RNN) to predict hyper-parameter and structure of DNN, consuming many resource even on Cifar-10. Project at ZJU may pay more attention to trade-off between model complexity and accuracy.

A Question about DRNet:

DRNet [3] models the statistical relations. The released code implements predicate recognition but I feel $\mathbf{q}'_\mathbf{r} = \sigma(\mathbf{W_r q_r} + \mathbf{W_{rs} q_s} + \mathbf{W_{ro} q_o})$ slight different with $\mathbf{q}'_\mathbf{r} = \sigma(\mathbf{W_r x_r} + \mathbf{W_{rs} q_s} + \mathbf{W_{ro} q_o})$ in the paper. I may missing some key points and the slide for my tentative talk can be found at github/seminar. Meanwhile, I guess implement a costumed inference unit mentioned in [3] and design network based on this unit may further boost performance.

Thanks a lot for your careful reading, I am looking forward to hearing from you!

Xinglu Wang
5-515 YuQuan Campus – Zhejiang University – Hangzhou, China
(+86)-17816872816 • *3140102282@zju.edu.cn*

References

[1] Wang L, Xiong Y, Wang Z, et al. Temporal segment networks: Towards good practices for deep action recognition[J]. arXiv preprint arXiv:1608.00859, 2016. Publishing Company , 1984-1986.

[2] Yang, Shuo, et al. "Face Detection through Scale-Friendly Deep Convolutional Networks." arXiv preprint arXiv:1706.02863 (2017).

[3] Dai, Bo, Yuqi Zhang, and Dahua Lin. "Detecting Visual Relationships with Deep Relational Networks." arXiv preprint arXiv:1704.03114 (2017).

[4] Zhang, Xingcheng, et al. "Polynet: A pursuit of structural diversity in very deep networks." arXiv preprint arXiv:1611.05725 (2016).

[5] Dai, Jifeng, et al. "Deformable Convolutional Networks." arXiv preprint arXiv:1703.06211 (2017).

[6] Zoph B, Le Q V. Neural architecture search with reinforcement learning[J]. arXiv preprint arXiv:1611.01578, 2016.