

论文解读之 Learning to Steer by Mimicking Features from Heterogeneous Auxiliary Networks

会议：AAAI 2019

原文地址：http://personal.ie.cuhk.edu.hk/~ccloy/files/aaai_2019_learning.pdf

代码：<https://github.com/cardwing/Codes-for-Steering-Control>

本文[1]主要提出一种基于多层异构辅助网络特征模仿(multi-level heterogeneous auxiliary network feature mimicking)的网络训练方法,做的任务是无人驾驶汽车的方向盘控制(steering control)/方向盘转角预测(steering angle prediction)。

论文内容简介:

无人驾驶任务通常被拆分成一系列的子问题,包括感知,决策,路径规划和控制[2]。最近的方法[3]通常端到端(end-to-end)训练一个卷积神经网络(convolutional neural network)来完成无人驾驶任务。这个卷积网络的输入是图片或视频,输出是汽车方向盘转角(steering angle)。

汽车方向盘转角通常作为这些网络训练的唯一监督信号[3, 4]。一些研究[5, 6]通过使用多任务学习(multi-task learning)来提升模型的预测准确度。比如,网络同时预测汽车的速度和方向盘转矩(steering torque)并用相应的标签(label)来监督。这些额外的监督信号虽然包含一定的有用信息,但是却无法保证网络进行有效的特征学习(representation learning)以很好地利用丰富的周围环境信息(environmental context),例如路面位置结构信息和物体车辆的位置信息。这些信息对于无人驾驶任务至关重要。无法捕捉并有效利用这样的空间信息使得现有方法往往在一些比较困难的驾驶场景(例如强烈的光线变化,强阴影,急转弯或交通状况拥挤)下表现不佳。

一个可行的提升模型环境信息捕捉利用能力的方法就是将多任务学习从速度转矩预测扩展到更加复杂的任务,例如语义分割(semantic segmentation),车道线检测(lane detection)或者光流预测(optical flow estimation)。这些任务可以捕捉利用环境的空间结构信息以及物体的移动信息(motion)进而提升模型的方向盘转角预测精度。不过,引入这些辅助任务需要额外任务相关的标注,而这些标注往往费时昂贵。一个可选择的方法就是在相关任务上预训练方向盘转角预测模型,然后再在汽车方向盘转角数据集上微调(finertune)。这个方法的确可以不需要昂贵的额外标注,不过在实验中,我们发现这个方法只能很有限地提高模型方向盘转角预测的精度。

于是,我们提出了基于多层异构辅助网络特征模仿的网络训练方法,这种训练方法不需要额外的人工标注信息。具体来说,我们将训练好的 PSPNet[7](语义分割模型)和 FlowNet2[8](光流预测模型)在目标数据上生成特征(feature),然后再通过知识蒸馏(knowledge distillation)让我们的模型去模仿这些特征。我们之所以这么做是基于一个有趣的实验现象:当我们把由 PSPNet 和 FlowNet2 生成的特征先降维再显示到二维空间时,我们发现这些特征和方向盘转角相关性很高。因此,我们想到通过学习这些辅助网络的特征来提升模型的特征学习(representation learning)能力,进而提高模型的预测精度。最后,借助我们提出的网络训练方法,我们的方向盘转角预测模型在三个数据集(Udacity[9], Comma.ai[10] 和 BDD100K[11])上取得了最好的结果。(详细内容见原文)

References:

- [1] Hou, Y., Ma, Z., Liu, C. and Loy, C.C., 2018. Learning to Steer by Mimicking Features from Heterogeneous Auxiliary Networks. *arXiv preprint arXiv:1811.02759*.
- [2] Paden, B., Čáp, M., Yong, S.Z., Yershov, D. and Frazzoli, E., 2016. A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Transactions on intelligent vehicles*, 1(1), pp.33-55.
- [3] Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J. and Zhang, X., 2016. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.
- [4] Pomerleau, D.A., 1989. Alvin: An autonomous land vehicle in a neural network. In *Advances in neural information processing systems* (pp. 305-313).
- [5] Chowdhuri, S., Pankaj, T. and Zipser, K., 2017. Multi-modal multi-task deep learning for autonomous driving. *arXiv preprint arXiv:1709.05581*.
- [6] Yang, Z., Zhang, Y., Yu, J., Cai, J. and Luo, J., 2018, August. End-to-end Multi-Modal Multi-Task Vehicle Control for Self-Driving Cars with Visual Perceptions. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 2289-2294). IEEE.
- [7] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J., 2017. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881-2890).
- [8] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A. and Brox, T., 2017. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2462-2470).
- [9] Udacity. 2018. <https://github.com/udacity/self-driving-car/>. Accessed: 2018-02-20.
- [10] Santana, E. and Hotz, G., 2016. Learning a driving simulator. *arXiv preprint arXiv:1608.01230*.
- [11] Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V. and Darrell, T., 2018. BDD100K: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*.