计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2019,28(1):239−244 [doi: 10.15888/j.cnki.csa.006730]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 NVIDIA Jetson TX2 的道路场景分割①

李诗菁, 卿粼波, 何小海, 韩　杰

(四川大学 电子信息学院, 成都 610065)
通讯作者: 卿粼波, E-mail: qing_lb@scu.edu.cn

摘　要: 图像语义分割是计算机视觉领域重要研究方向之一, 其中基于深度学习的语义分割相较于传统分割算法更为高效可靠, 可应用于交通监控、自动驾驶等领域的场景理解阶段. 但复杂的分割网络在嵌入式平台上的推理速度较低, 难以进行实际应用. 因此针对交通监控、无人驾驶等应用背景, 在嵌入式平台 NVIDIA Jetson TX2 上, 采用基于深度卷积编解码器结构的图像分割网络, 对道路场景进行语义分割, 并基于 NVIDIA 的推理加速器 TensorRT2, 完成网络模型简化、网络自定义层添加与 CUDA 并行优化, 实现了对网络推理阶段的加速. 实验结果表明, 加速引擎在 TX2 上的推理速度约为原模型的 10 倍, 为复杂分割网络在嵌入式平台上的应用提供了支持.

关键词: 场景理解; 深度学习; Tensor RT2 语义分割; NVIDIA Jetson TX2

引用格式: 李诗菁,卿粼波,何小海,韩杰.基于 NVIDIA Jetson TX2 的道路场景分割.计算机系统应用,2019,28(1):239–244. http://www.c-s-a.org.cn/1003-3254/6730.html

Road Scene Segmentation Based on NVIDIA Jetson TX2

LI Shi-Jing, QING Lin-Bo, HE Xiao-Hai, HAN Jie

(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Image semantic segmentation is one of the most important research directions of computer vision. Compared with traditional algorithms, image segmentation based on deep-learning performs better, and can be applied to the scene understanding stage of traffic monitoring and automatic drive. However, the speed of complex segmentation network on embedded platform is too low to be practically applied. Therefore, in view of the application of traffic monitoring and automatic drive, the image segmentation network based on deep convolutional encoder-decoder architecture was used to complete the road scene segmentation on the embedded platform NVIDIA Jetson TX2. Meanwhile, in order to accelerate the network, the model was simplified and transformed to engine based on TensorRT2 provided by NVIDIA, which including plugin layers adding and CUDA parallel optimization. The experimental results show that the speed-up ratio can reach ten, which provides support for the application of the complex structure segmentation network on the embedded platform.

Key words: scene understanding; deep-learning; Tensor RT2 semantic segmentation; NVIDIA Jetson TX2

近年, 深度学习的迅猛崛起给各个科技领域的发展带来了巨大影响. 在计算机视觉领域, 图像语义分割是重要的基础研究问题之一. 传统的图像分割一般采用基于阈值、边缘、区域的分割方法, 根据颜色纹理等人工标定的特征完成分割, 过程复杂且局限性较大[1]. 基于深度学习的图像语义分割网络的分割性能远

① 基金项目: 成都市科技项目 (2016-XT00-00015-GX); 四川省教育厅科研项目 (18ZB0355)
Foundation item: Science and Technology Program of Chengdu Municipality (2016-XT00-00015-GX); Science and Technology Research Program of Education Bureau, Sichuan Province (18ZB0355)
收稿时间: 2018-07-17; 修改时间: 2018-08-09; 采用时间: 2018-08-16; csa 在线出版时间: 2018-12-26

超传统算法, 可应用到诸多领域如无人驾驶、智能安防、交通监控、机器人中[2]. 在较有代表性的图像分割网络 FCN (Fully Convolutional Networks)[3]、SegNet (SegNet: A Deep ConvolutionalEncoder-Decoder Architecture for ImageSegmentation)[4]、PSPNet (Pyramid Scene Parsing Network)[5]和 ENet(ENet: A Deep Neural Network Architecture forReal-Time Semantic Segmentation) 中, FCN 是最早的图像分割网络之一, 准确度较低且耗时长; PSPNet 分割的像素精确度高, 但耗时较长; ENet 分割速度较快, 但分割效果较不理想; SegNet 兼具较高的分割准确度与速度, 是完成道路场景分割任务较好的选择.

随着人工智能的不断发展, 深度学习网络在嵌入式平台上部署的需求也日益增长. 目前国内外推出的可部署深度学习的主流芯片一般分为 FPGA 芯片和 GPU 芯片两类. 在 FPGA 上已实现了一些小规模的神经网络, 如 Zhang 等人在 Xilinx Virtex 7485T FPGA 上实现了 AlexNet[6], 卢冶、陈瑶等人在 XC7Z020 FPGA 上实现了 LeNet-5、Cifarnet 网络[7]. 基于 FPGA 的深度学习应用的功耗小于 GPU, 但受限于 FPGA 开发难度与硬件资源, 目前大规模复杂结构的网络在嵌入式上的部署还是更多地基于 GPU 实现. 已有许多深度学习网络逐渐被应用到了嵌入式 ARM+GPU 平台如 NVIDIA JetsonTX2 (简称 TX2) 上, 构建了目标检测、感知导航等系统[8,9]. 因分割网络相较于图像分类、目标检测更难实现, 大部分高准确度的分割网络在低功耗的嵌入 GPU 平台上运行速度极慢, 而较少地被应用到嵌入式平台上.

因此, 针对无人驾驶与交通监控等应用背景, 本文选取 SegNet 网络在嵌入式平台 TX2 上实现道路场景理解, 并采用模型简化与转换网络模型为加速引擎的方式, 在基本保持原网络分割准确度的情况下大幅度提升了分割网络在嵌入式平台中的推理速度. 本文首先合并 BN 层简化了网络模型, 然后基于英伟达的 TensorRT2 前向推理加速器采用水平层集成, 关联层消除, 权值精度减半等优化措施, 将网络模型转换成了加速引擎进行前向推理. 下面将分别就网络结构与模型简化、基于 TensorRT2 的加速引擎构建两部分进行介绍.

1 网络结构与网络模型简化

1.1 网络结构

本文采用 SegNet 网络, 实现将道路场景分为行人、道路、天空、标志、车辆、交通标线、杆状物、自行车、侧路、植物、建筑、围墙共 12 类的任务.

SegNet 网络基本结构为自动编-解码器结构, 采用 VGG16 前 13 层卷积层作为编码器, 后接 13 层解码网络与一个分类层. SegNet 关键在于存储了编码网络中每个的池化层中的最大值与其空间信息, 用于对应解码器的非线性上采样, 这极大精确了分割中的边界定位, 减少了编码器到解码器的参数量, 使得 SegNet 在速度与内存利用上都具有很大优势.

1.2 网络模型简化

在将网络模型转换为加速引擎之前, 本文对 SegNet 网络进行了模型简化. SegNet 网络中引入了 BN 层, 这种层可在训练时期加速网络收敛, 但在推理过程中会增大内存消耗, 降低推理速度. 可将 BN 层与其相连的卷积层合并以简化模型加速推理, 如图 1 所示.

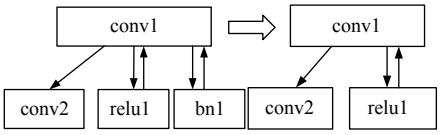

图 1 BN 层合并

卷积层与 BN 层都是线性变换, 提供了合并的可能性. 设输入的 mini-batch: $\{x_{1...n}\}$, BN 层变化如式 (1):

$$y_i \leftarrow \gamma \frac{x_i - E(x)}{\sqrt{\text{var}(x)+\varepsilon}} + \beta \tag{1}$$

式 (1) 中, $i \in [0,n]$, 且i为整数, γ与β是训练中得到的参数, ε为趋于 0 的常数.

在推理阶段, 原卷积层权值为w, 偏置项为b, 将 BN 层合并到卷积层, 得到卷积层新的权值为 $w_{\text{new}} = w*\gamma$, 新的偏置值为$b_{\text{new}} = b*\gamma+\beta$, 故可构成新的卷积层, 得到简化后的网络模型. BN 层合并之后可以与卷积层共用 Blob 数据, 从而可以减少内存占用, 有利于速度提升.

2 基于 TensorRT2 的加速引擎构建

SegNet 在 TX2 上的推理速度极慢, 因此本文采用 NVIDIA 推出的 TensorRT2 对网络模型进行加速.

2.1 基本流程

TensorRT 是 NVIDIA 公司推出的深度学习网络推理引擎, 可优化已有网络模型, 大幅提升神经网络在

如机器人、自动驾驶平台上的推理速度. 目前 NVIDIA 公司已推出四个版本的 TensorRT. 不同版本 TensorRT 可支持的深度学习框架不同.

因为本文使用的 SegNet 网络采用 caffe 架构, 且需要插入 TensorRT 不支持的层, TensorRT2 版本已经能满足本文需求. 故本文使用 TensorRT2, 基于训练后的网络模型生成加速推理引擎, 完成高效的 GPU 推理性能优化. 加速后的模型无需深度学习架构支持, 对平台的依赖性极小, 无论在高性能 GPU(如 NVIDIA Titan X) 上, 还是在嵌入式平台 (例如 NVIDIA TX2) 上都拥有不俗的加速能力. 加速引擎构建的基本流程如图 2.

图 2 加速引擎构建流程

构建过程中首先创建网络, 导入网络模型并利用 NvcaffeParser 进行解析, 然后根据网络结构与定义采取优化措施生成加速引擎. 生成的加速引擎可序列化存储到磁盘中. 在执行阶段反序列化加速引擎, 进行输入与输出口的绑定、GPU 显存分配与计算内核启动, 执行加速引擎得到分割结果.

2.2 自定义层添加

在上文的网络模型解析过程中, 需将原网络模型中的层转换为其支持的形式. 加速引擎中除部分常见层如卷积层、激活层、全连接层等可直接转换外, 其它 TensorRT 不支持的层需自行定义, 下文将这些层统称为自定义层. TensorRT 中支持的层可直接由上文提到的 NvcaffeParser 进行解析, 自定义层则需采用 plugin 接口添加入 NvcaffeParser 中.

自定义层 plugin 添加的主要流程如下: 输出确定、层配置、工作空间分配、资源管理、序列化与层执行. 在加速引擎中, 一个层的输入或输出定义为 tensor, tensor 具有数据类型与三个维度分量, 即通道数 C、宽度 W 与高度 H. 在开始插入层时, 需对插入层的输出进行定义, 需确定插入层的输出数目与输出 tensor 的三个维度分量$\{C, H, W\}$; 完成输出确定后, 将进行层配置, 该过程主要获取输入 tensor 的形式; 在层配置后, 加速引擎会分配临时的工作空间在自定义层之间共享以达到内存利用率最大化; 同时还需要进行资源管理的配置, 主要是通过层资源初始化与销毁来完成资源分配与释放; 在进行完成上述部分后, 需判断加速引擎是否需要序列化到磁盘中, 若需序列化, 则将自定义层的参数与网络的其余部分合并以便后续整个网络的序列化存储; 最后在层执行阶段, 主要完成层的算法实现, 如果未选择序列化存储, 则直接在资源分配后执行该过程, 若已进行序列化存储, 则提取序列化参数后执行该过程, 整体过程如图 3 所示.

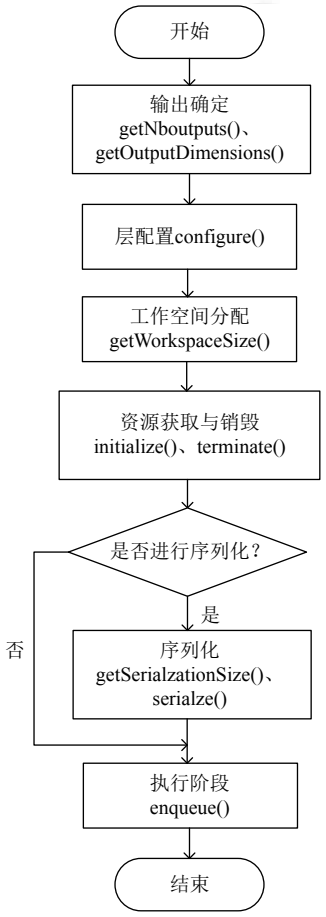

图 3 plugin 层插入流程

本文将网络中需两个输出口的 pooling 层、upsample 层与 argmax 分类层添加到加速模型中, 这些添加的层是网络的关键结构, 如图 4 所示. 本文将 pooling 层分出了两个输出口, 将池化结果送入下一级编码器的同时将池化的最大特征值与位置信息输入了对应 upsample 层中. upsample 层将接收的池化最大特征值按照位置信息填入上采样的稀疏特征图中, 空缺位置补零, 作为解码器的输入, 如图 5. 最终经过层层解码后的特征图输入 argmax 层, 得到每个像素最大可能性的分类结果.

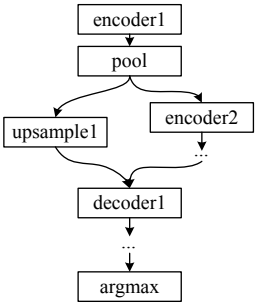

图 4 plugin 层结构

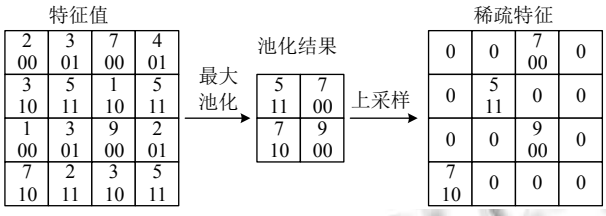

图 5 plugin 层实现过程

2.3 优化措施

本文的优化措施主要分为图 1 中的引擎生成阶段的优化与图 3 自定义层添加中执行阶段 enqueue( ) 时的 CUDA 编程优化.

网络模型中的各个层经过解析之后, 在引擎生成时主要采用了以下措施进行优化: 清除未使用的输出层以避免不必要的计算; 整合垂直结构的卷积层、ReLU 激活层、Bias 偏置项为 CBR 层; 将拥有同一输入、相同操作、相似参数和不同输出的水平层集成为一个层, 例如图 6; 在平台支持的情况下, 将数据精度需求从 32 位 float 降低为半精度的 fp16, 以此提升计算效率.

在自定义层添加时执行层算法的函数 enqueue( ) 中, 本文采用了 CUDA 编程的方式将算法中大量耗时的计算分配到 GPU 上并行实现. 原网络最后的分类层 argmax 层为 CPU 实现, 若同样采用 CPU 实现, 需将输入层的数据从 GPU 上复制到 CPU 上. 数据复制将耗费大量时间, 且 argmax 层算法在 CPU 上的执行时间也过长. 因此本文在 GPU 上实现了 argmax 层前馈算法. 在 CUDA C 中的 host 函数只可由 CPU 调用, device 函数则只可运行在 GPU 上, 且不能调用常用函数. 因此 argmax 层前馈算法中调用的部分算法, 并不能在 GPU 上调用. 为此本文针对网络的实际应用简化了相应算法以便 GPU 实现. 最终在得到每个像素值 12 类分类结果后, 在一维索引空间中并行 360*480 个线程, 将位置一一对应的特征值的 12 个分类结果进行排序, 得到整幅图像像素级别的分类.

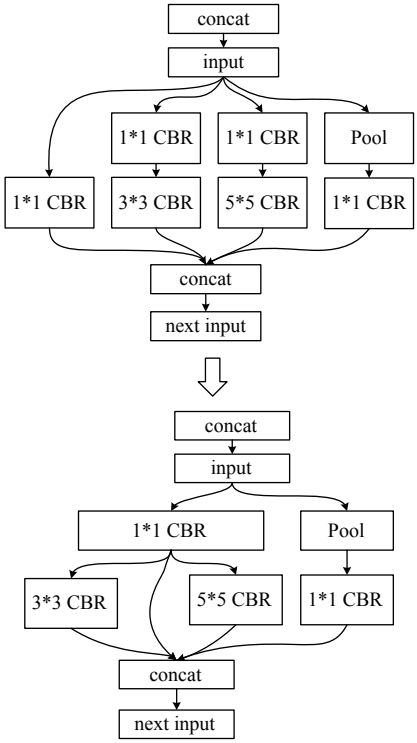

图 6 水平层集成

3 实验测试

3.1 网络训练

本文采用 GPU 模式进行网络训练, 硬件为基于 Pascal 架构的 NVIDIA Titan X 显卡, 该卡显存为 12 GB.

为满足针对交通监控与自动驾驶的道路场景理解的目的, 本文从包含城市交通场景的数据库 KITTI[10]、Camvid[11]与 Cityscapes[12]中选取了 1300 张图像作为训练数据集, 并将原图与标签裁剪为统一尺寸. 此外, 选取的网络训练需要单通道标签, 故将原数据集中 RGB 标签转换为单通道, 并将每个像素转换为代表天空、行人、围墙、植物、交通标志、侧道、道路、车辆等 12 个分类的灰度值. 在训练时, 为提高分类准确度, 使用预先训练好的 VGG16 网络模型的权值对 SegNet 编码网络的权值进行初始化. 在 20 000 次 iteration 后, 网络准确度就已经达到 80% 以上, 在迭代 36 000 次时, 达最高准确度 92%, 迭代 38 000 次时达到最低损失 0.019.

3.2 测试结果

在训练得到最优网络模型的基础上, 本文完成了加速引擎 (下文称为 SegEng) 的构建, 并将其序列化后存储到磁盘上. NVIDIA TX2 平台支持 fp16, 故本文生

成 SegEng 时采用了半精度模式, 将其大小缩减为 SegNet 模型的一半. 在 NVIDIA TX2 上, 本文对 GPU 模式下使用 cudnn6.0 加速的 SegNet 与 SegEng 的前向推理性能进行了测试对比.

(a) 输入图像 (b) SegNet分割 (c) SegEng分割

图 7 分割结果

分割图对比如图 7 所示. 在图像分割领域, 常用以下三个指标衡量分割精度: 像素精确度 (*G*) 表示图像中被正确分类的像素比; 类平均精确度 (*C*) 表示所有类被正确分类的平均值; 交除并均值 (Mean Intersection over Union, MIoU) 代表所有类的 IoU 平均值.

本文从测试集中选取图片, 使用上述三个指标对迭代了 40 000 次得到的 SegNet 模型与 SegEng 的分割结果进行评估, 得到表 1.

表 1 分割精度对比 (单位: %)

模型	*G*	*C*	*MIoU*
SegNet	85.4	68.9	51.7
SegEng	85.4	68.9	51.7

原始模型中的权值精度为 32 位, 加速引擎舍弃了权值中后 16 位的冗余信息, 将权值精度降低为 16 位, 权值中丢弃的后 16 位信息对最终的像素级别分类概率影响极小, 由表 1 中可看到并未影响到最终的分割准确度.

在 NVIDIA TX2 上, 本文对不同分辨率的 1000 帧视频进行了 10 次测试后取平均值, 得到了表 2 中 SegNet 与 SegEng 推理时间的对比.

表 2 单帧图像推理时间对比 (单位: ms)

模型	480*320	640*360	1280*720
SegNet	802.38	1215.07	5114.80
SegEng	88.14	112.13	489.36
加速比	9.10	10.84	10.45

在经过上文的层合并、网络层优化、CUDA 加速与 TensorRT 内部的卷积优化、内存优化、精度减半等优化措施后, 由表 2 可看出, 在 TX2 上, 相较于 SegNet, 本文生成的 SegEng 达到了约十倍的加速比, 大幅度提高了 SegNet 在嵌入式平台上的推理速度. 在分割 480*320 大小的图片时, 分割速度可达 10 fps, 具有在实际道路场景解析的应用潜力.

4 结语

本文针对无人驾驶与交通监控的应用背景, 采用了图像语义分割网络 SegNet 完成将道路场景分割为行人、车辆、道路、植物、侧道、交通标志、自行车等 12 类对象的任务. 同时本文为提升分割网络在嵌入式平台上的推理速度, 对网络模型进行了简化, 然后基于 NVIDIA 推出的 TensorRT2, 采用集成水平层、消除多余输出层、采用权值精度减半、CUDA 并行编程等优化措施, 完成了网络模型加速并生成了加速引擎. 在 NVIDIA TX2 嵌入式平台上, 本文生成的加速引擎无需深度学习架构支持, 在分割精度基本无影响情况下, 推理速度可达原网络的十倍, 为复杂结构分割网络在嵌入式平台上的应用提供了支持.

参考文献

1 陈鸿翔. 基于卷积神经网络的图像语义分割[硕士学位论文]. 杭州: 浙江大学, 2016. 8–10.

2 吴宗胜, 傅卫平, 韩改宁. 基于深度卷积神经网络的道路场景理解. 计算机工程与应用, 2017, 53(22): 8–15. [doi: 10.3778/j.issn.1002-8331.1708-0195]

3 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 3431–3440.

4 Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495. [doi: 10.1109/TPAMI.2016.2644615]

5 Zhao HS, Shi JP, Qi XJ, *et al*. Pyramid scene parsing

network. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 6230–6239.

6 Zhang C, Li P, Sun GY, *et al*. Optimizing FPGA-based accelerator design for deep convolutional neural networks. Proceedings of 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays. Monterey, CA, USA. 2015: 161–170.

7 卢冶, 陈瑶, 李涛, 等. 面向边缘计算的嵌入式 FPGA 卷积神经网络构建方法. 计算机研究与发展, 2018, 55(3): 551–562.

8 Tijtgat N, Van Ranst W, Volckaert B, *et al*. Embedded real-time object detection for a UAV warning system. Proceedings of 2017 IEEE International Conference on Computer Vision Workshops. Venice, Italy. 2017. 2110–2118.

9 Hui XL, Bian J, Yu YJ, *et al*. A novel autonomous navigation approach for UAV power line inspection. Proceedings of 2017 IEEE International Conference on Robotics and Biomimetics. Macau, China. 2017. 634–639.

10 Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA. 2012. 3354–3361.

11 Brostow G J, Fauqueur J, Cipolla R. Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters, 2009, 30(2): 88–97. [doi: 10.1016/j.patrec.2008.04.005]

12 Cordts M, Omran M, Ramos S, *et al*. The cityscapes dataset for semantic urban scene understanding. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 3213–3223.