ICANN19

28th International Conference on Artificial Neural Networks

Attentional Residual Dense Factorized Network for Real-time Semantic Segmentation

Lulu Yang [1,2,✉], Long Lan [2,3,✉], Xiang Zhang [2,3,✉], Xuhui Huang [4], and Zhigang Luo [1,2,✉]

[1]Science and Technology on Parallel and Distributed Laboratory, NUDT, Changsha 410073, China

[2]College of Computer, NUDT, Changsha 410073, China

[3]Department of Computer Science and Technology, NUDT, Changsha, Hunan, P.R. China, 410073

Emai:windychanchan@163.com, {long.lan,zhangxiang08,zgluo}@nudt.edu.cn

WeChat ID:Windy_YX

1. Topic and Motivations

Background: Semantic segmentation is in efforts to assign each pixel of an image a class label, e.g., car or road. It plays a core role in autonomous driving, which requires a scene understanding system that can operate in real time. Despite much progress in this regard, it is still a challenging task for the sake of difficulty in balancing between precision and efficiency.

Motivation: Although some work have conducted preliminary research on lightweight architecture networks, pursuing the state-of-the-art performance with a good trade-off between precision and efficacy still remains an open research issue for the task of real-time semantic segmentation.

2. AttRDFNet

1. The overall architecture of the AttRDFNet is shown in Fig.1. We design a residual dense factorized convolution block (RDFB) as shown in Fig.2, which reaps the benefits of low-level and high-level layer-wise features through dense connections to boost segmentation precision whilst enjoying efficient computation by factorizing large convolution kernel into the product of two smaller kernels. This reduces computational burdens and makes real time become possible.

2. To further leverage layer-wise features, we explore the graininess-aware channel and spatial attention modules to model different levels of salient features of interest, as shown in Fig.3.

3. AttRDFNet can run with the inputs of the resolution 512×1024 at the speed of 55.6 frames per second on a single Titan X GPU with solid 68.5% Mean IOU on the test set of Cityscapes. Experiments on the Cityscapes dataset show that AttRDFNet has real-time inference whilst achieving competitive precision against well-behaved counterparts.

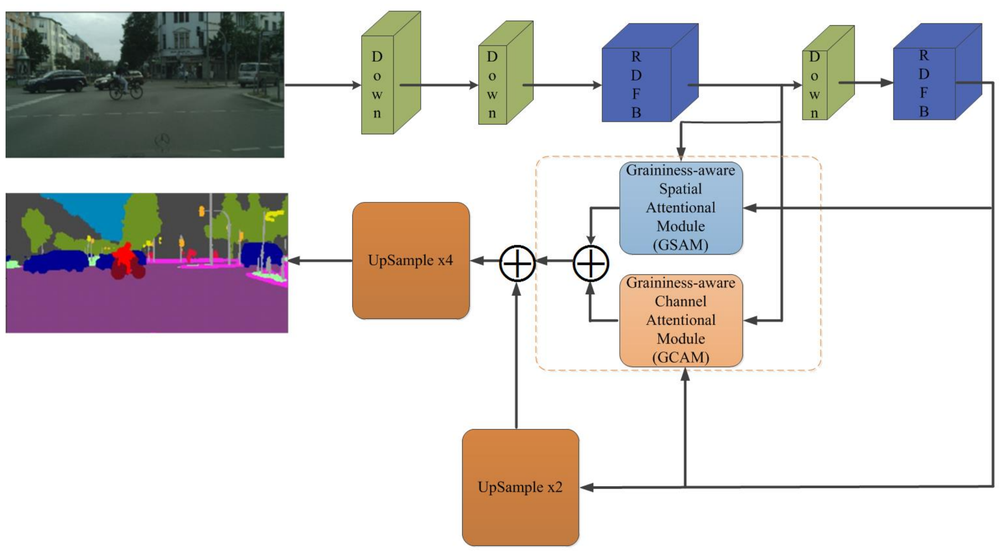

Fig.1. An overview of AttRDFNet.

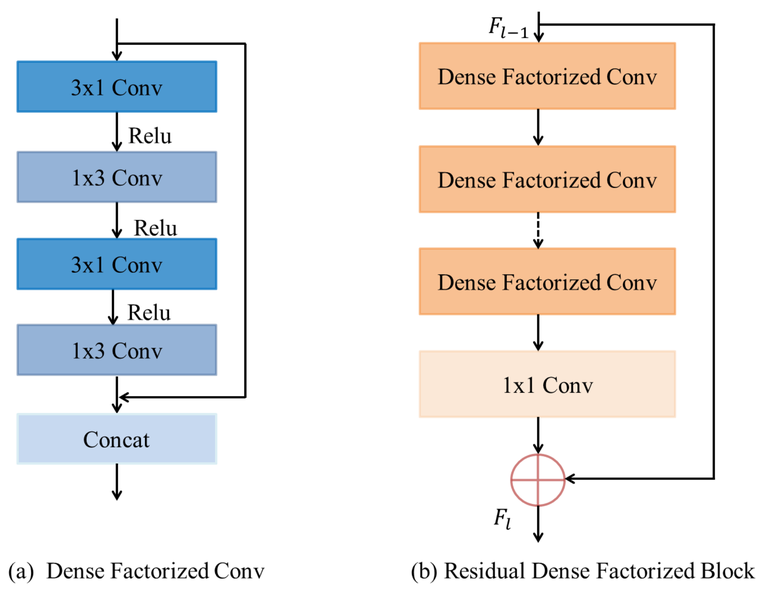

Fig.2. Dense Factorized Conv and Residual Dense Factorized Block.

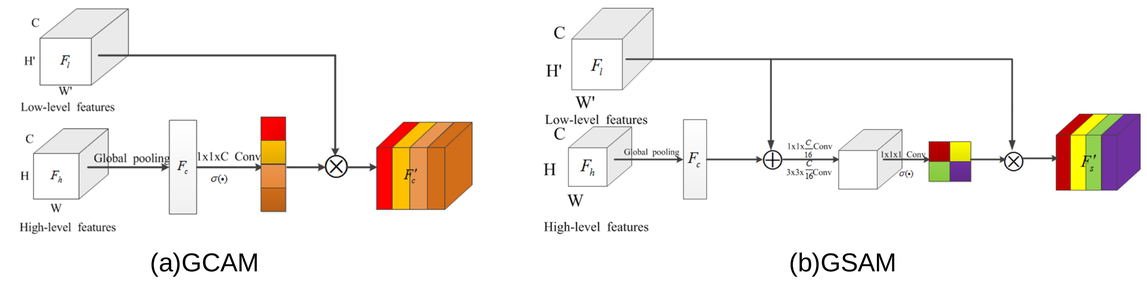

Fig.3. The structure of graininess-aware channel attentional module (GCAM) and graininess-aware spatial attentional module (GSAM).

3. Experimental Results

We compare our network with several well-established real-time segmentation networks on Cityscapes dataset. As shown in Table 1. The visualization results are shown in Fig.4.

Table 1. Comparison of the accuracy and speed of different methods on the Cityscapes test set, we evaluate on NVIDIA Titan X GPU with 1024×512 resolution input, n/a means no corresponding result is given. (Methods with * are pre-trained on ImageNet)

Model	Class mIoU(%)	Class iIoU(%)	Category IoU(%)	Category iIoU(%)	FPS
SegNet*	56.1	34.2	79.8	66.4	n/a
SQ*	59.8	32.3	84.3	66.0	n/a
DeepLabv3+*	82.1	62.4	92.0	81.9	n/a
FCN-8s*	65.3	41.7	85.7	70.1	n/a
Deeplab*	63.1	34.5	81.2	58.7	0.3
ERFNet	68.0	40.4	86.5	70.4	41.7
ContextNet	66.1	36.8	82.8	64.3	65.5
ENet	58.3	34.4	80.4	64.0	76.9
ESPNet*	60.3	31.8	82.2	63.1	112.9
Ours	68.5	40.0	86.0	70.4	55.6

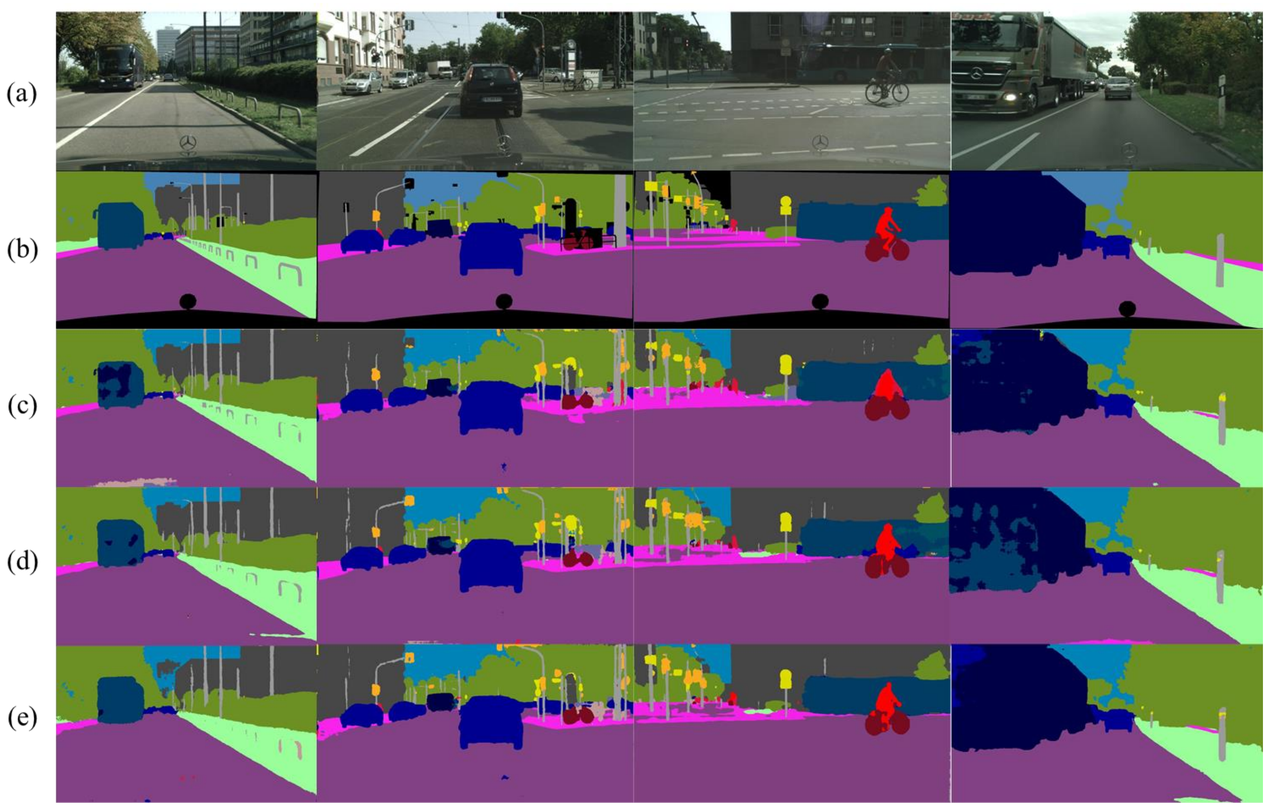

Fig.4. Some segmentation examples of the compared methods on Cityscapes validation dataset. From top to bottom, (a) input, (b) ground truth, (c) ENet [19], (d) ERFNet [21] and (e) AttRDFNet.

ManuscriptID:136