Tree-structured Kronecker Convolutional Networks for Semantic Segmentation

Tianyi Wu[1,2], Sheng Tang[1], Rui Zhang[1,2], Jintao Li[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.
[2] University of Chinese Academy of Sciences, Beijing, China.
{wutianyi, ts, zhangrui, jtli}@ict.ac.cn

https://github.com/wutianyiRosun/TKCN

Contribution

- We propose Kronecker convolutions, which can effec- tively capture partial detail information and enlarge the field of view simultaneously, without introducing extra parameters.
- We develop Tree-structured Feature Aggregation module to capture hierarchical contextual information and represent multi-scale objects, which is beneficial for better understanding complex scenes.
- Without any post-processing steps, our designed TKCN achieves impressive results on the benchmarks of PASCAL VOC 2012, Cityscapes and PASCAL- Context.2.

Architecture

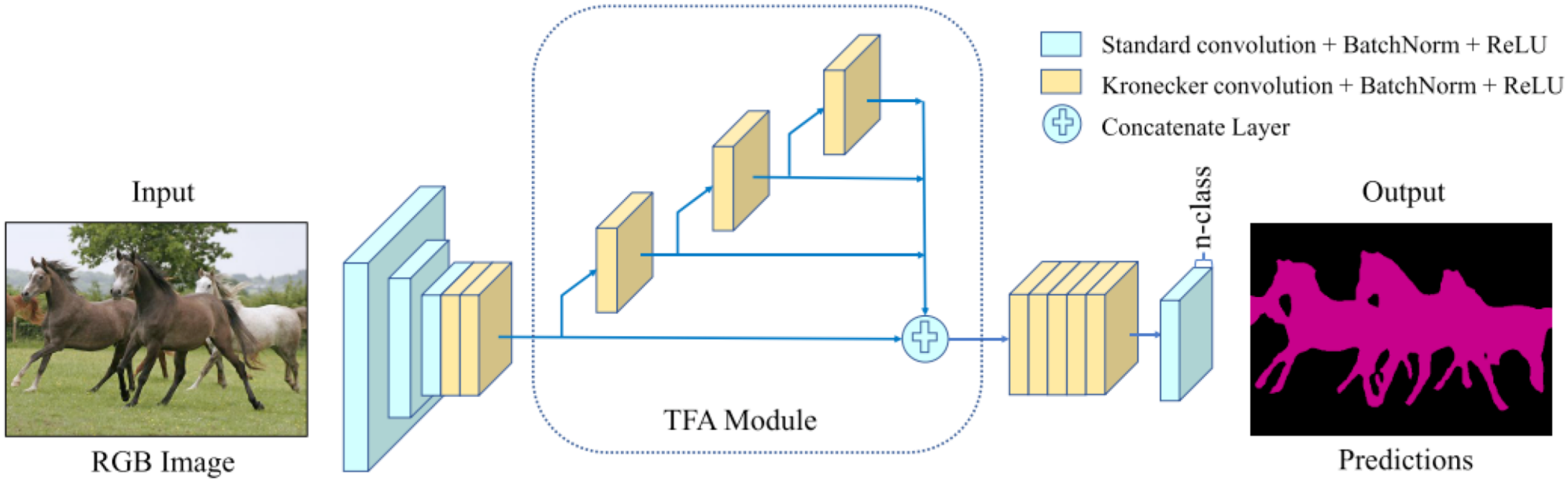


Figure 2. Architecture of the proposed TKCN. We employ Kronecker convolutions in ResNet-101 'Res4' and 'Res5'. Tree-structured Feature Aggregation module is implemented after the last layer of 'Res5'.

Kronecker Convolution

First of all, we provide a brief review of Kronecker product. If $\mathbf{A}$ is a $m \times n$ matrix and $\mathbf{B}$ is a $r \times s$ matrix, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $mr \times ns$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}. \quad (1)$$

$$K'(c_2, c_1) = K(c_2, c_1) \otimes F,$$

$$F = \begin{bmatrix} \mathbf{I}_{\mathbf{r_2} \times \mathbf{r_2}}, & \\ & \mathbf{O}_{(\mathbf{r_1}-\mathbf{r_2}) \times (\mathbf{r_1}-\mathbf{r_2})} \end{bmatrix},$$

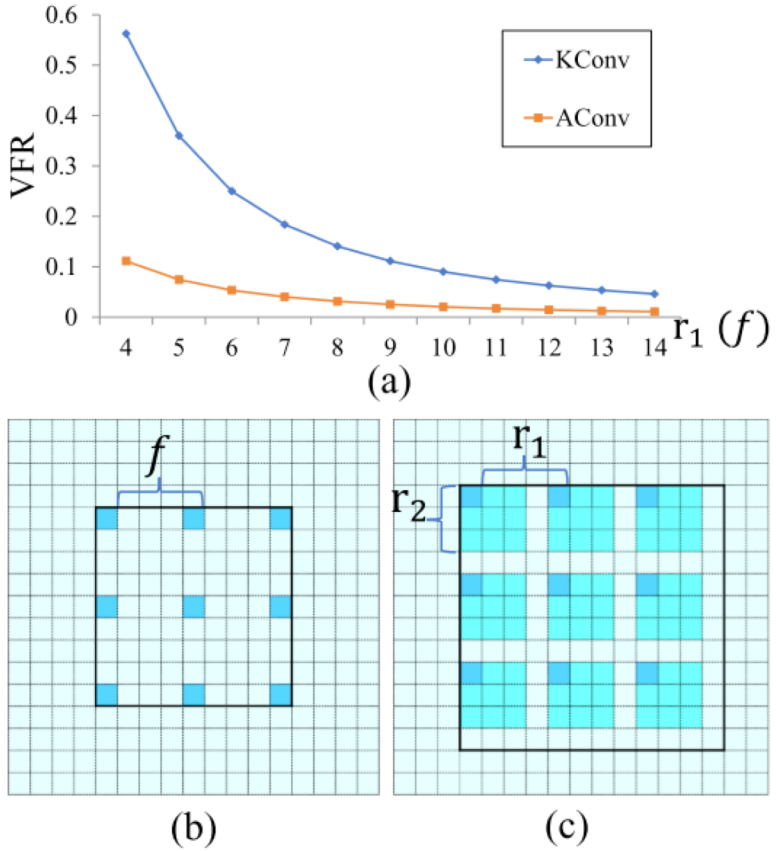


Figure 1. (a): Curves of the VFR for atrous convolutions (AConv) and Kronecker convolutions (KConv) with different rates. Inter-dilating factor r_1 of Kronecker convolution is equivalent to rate f of atrous convolution. Note that $r_2 = 3$. (b) Atrous convolution with rate $f = 4$. (c) Kronecker convolution with inter-dilating factor $r_1 = 4$, intra-sharing factor $r_2 = 3$. In (b) and (c), cells in the black boxes represent feature vectors in convolutional patches, while cyan and blue cells represent feature vectors involved in computation.

Result

Table 1. Evaluation results of Kronecker convolution (KConv) with different intra-sharing factor r_2 on PASCAL VOC 2012 validation set.

r_1	r_2	**mIoU(%)**	**Acc(%)**
6	1	77.03	94.97
6	3	78.37	95.25
6	5	78.75	95.36
10	1	78.01	95.17
10	3	78.53	95.24
10	5	78.93	95.34
10	7	79.50	95.53
10	9	79.71	95.54

Table 2. Comparison between Kronecker convolutions (KConv) and atrous convolutions (AConv) on PASCAL VOC 2012 validation set.

Method	r_1	r_2	**mIoU (%)**	**Acc (%)**
AConv (Baseline)	4	1	75.98	94.80
KConv	4	3	76.70	94.98
AConv	6	1	77.03	94.97
KConv	6	5	78.75	95.36
AConv	8	1	78.14	95.19
KConv	8	5	78.81	95.30
AConv	10	1	78.01	95.17
KConv	10	7	79.50	95.53
AConv	12	1	78.18	95.21
KConv	12	9	79.79	95.53

Table 3. Evaluation results of TFA module on PASCAL VOC 2012 validation set. **KConv:** employing Kronecker convolution on baseline model 'res4' and 'res5'.

Method	**mIoU (%)**	**Acc (%)**
Baseline (Baseline)	75.98	94.80
Baseline + TFA_S	80.18	95.56
Baseline + TFA_L	81.26	95.83
Baseline + KConv	76.70	94.98
Baseline + KConv + TFA_S	81.34	95.96
Baseline + KConv + TFA_L	82.85	96.26

TFA_S configured with small factors (r1, r2) = {(6, 3), (10, 7), (20, 15)}

TFA_L configured with large factors (r1, r2) = {(10, 7), (20, 15), (30, 25)}.

Comparison with State-of-the-Arts

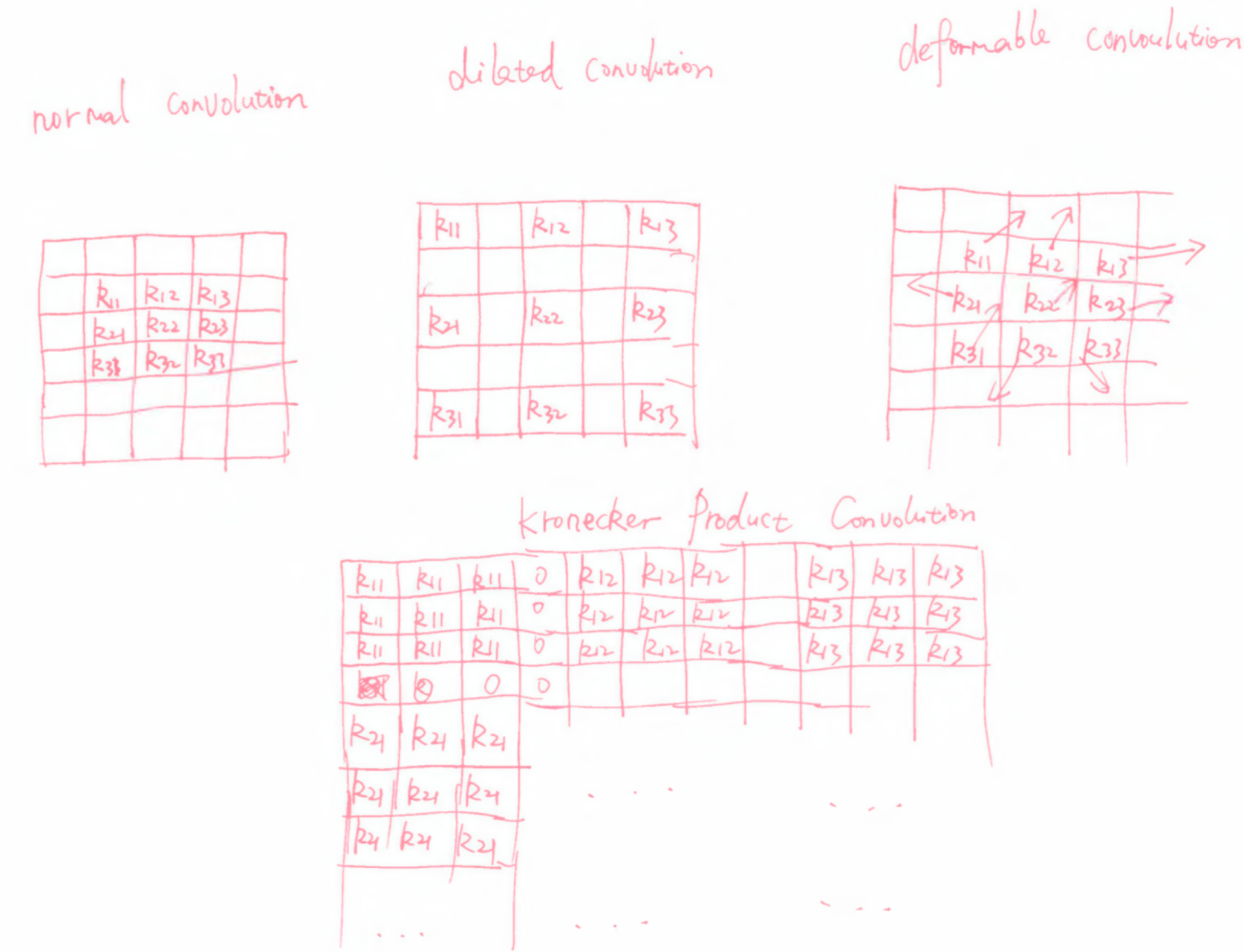
Table 4. Per-class mean intersection-over-union (IoU) results on the PASCAL VOC 2012 segmentation challenge test set, only using VOC 2012 for training. **Ms:** employing multi-scale inputs with average fusion during testing.

Method	mIoU (%)
FCN [29]	62.2
GCRF [32]	73.2
Piecewise [21]	75.3
DeepLab [2]	79.7
LC [19]	80.3
RAN-s [14]	80.5
RefineNet [20]	82.4
PSPNet_Ms [42]	82.6
DFN_Ms [37]	82.7
EncNet_Ms [39]	82.9
Deeplabv3+ [3]	89.0
TKCN	82.4
TKCN_Ms	83.2

Table 5. Per-class mean intersection-over-union (IoU) accuracy on Cityscapes test set, only training with the fine set. **Ms:** employing multi-scale inputs with average fusion during testing.

Method	mIoU (%)
CGNet [33]	64.8
FCN [29]	65.3
DeepLab [2]	70.4
LC [19]	71.1
RefineNet [20]	73.6
FoveaNet [18]	74.1
GRLRNet [40]	77.3
SAC_Ms [41]	78.1
PSPNet_Ms [42]	78.4
BiSENet_Ms [36]	78.9
DFN_Ms [37]	79.3
DenseASPP_Ms [35]	80.6
TKCN	78.9
TKCN_Ms	79.5

Summary


normal convolution
dilated convolution
deformable convolution
kronecker product Convolution