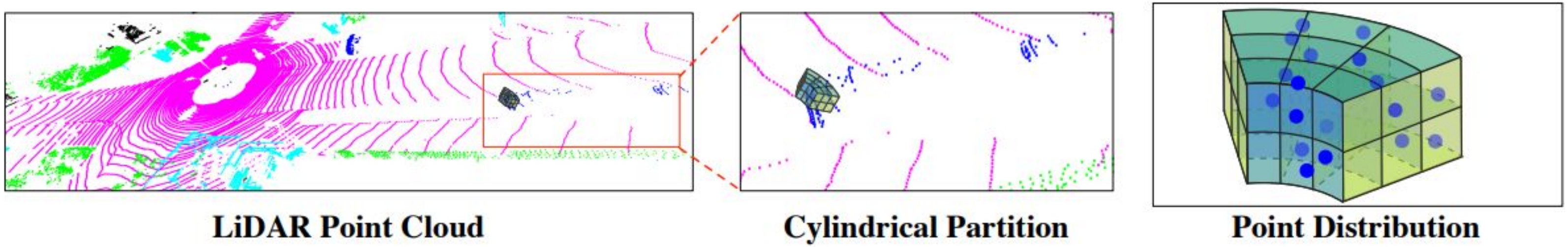
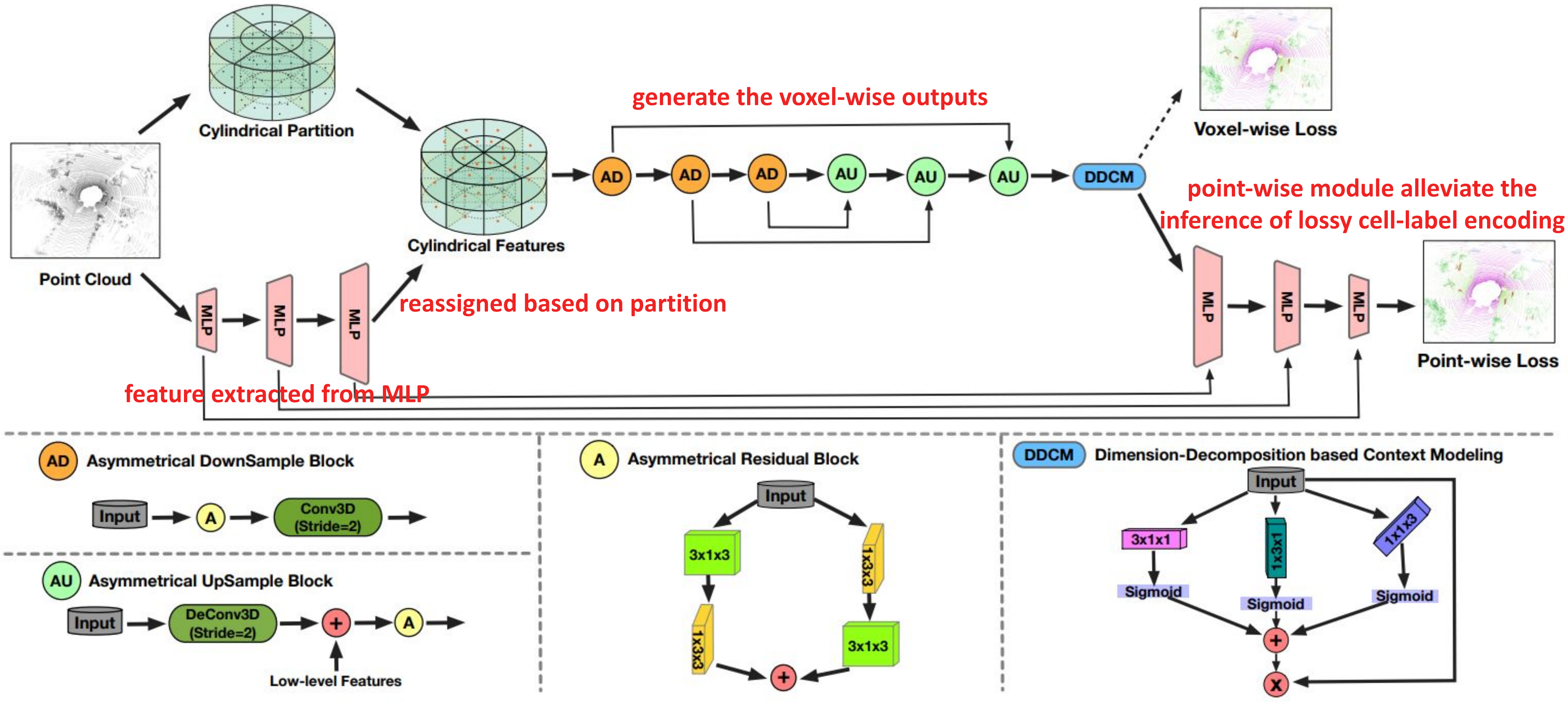
Paper Review

Cylindrical and Asymmetrical 3D Convolution Networks for LiDAR Segmentation

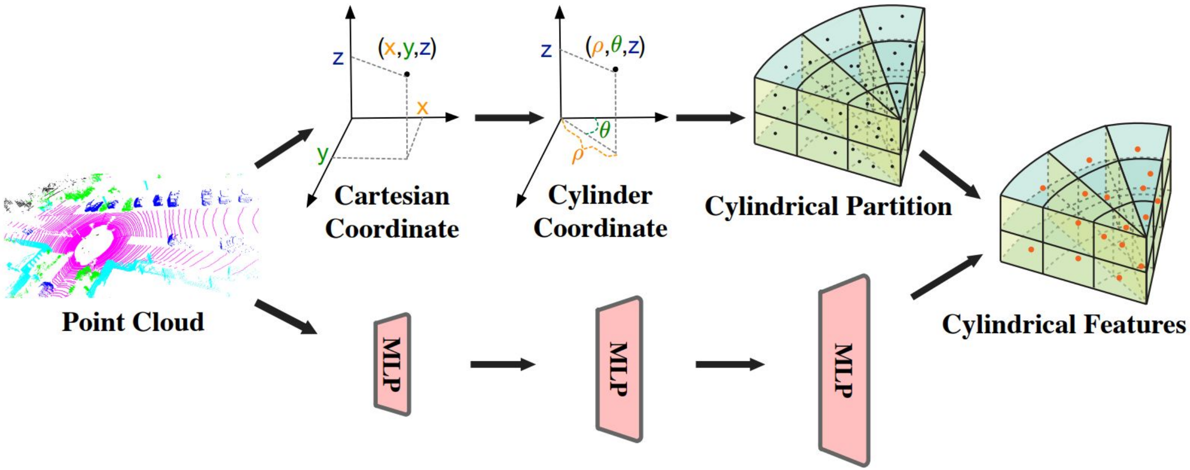
CVPR, 2021



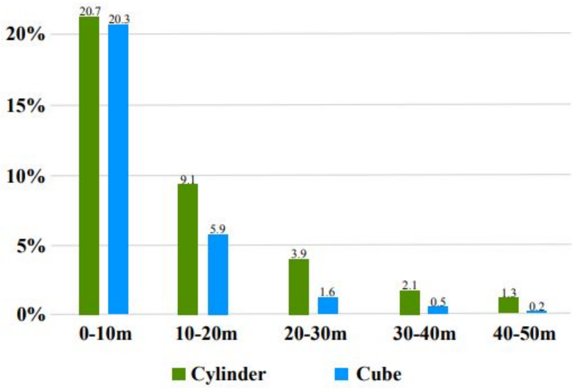
Architecture



Cylindrical Partition



Proportion of non-empty cells



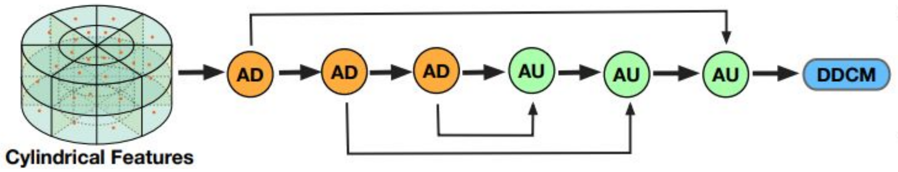
: Generate the more balanced point distribution

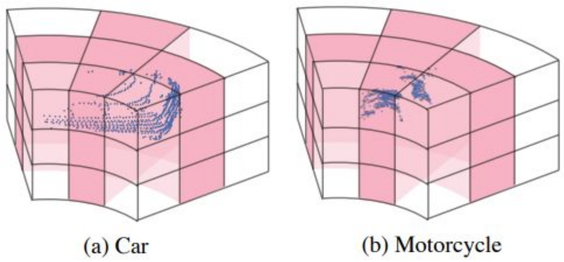
- **Increasing grid size** to cover the father-away region

- **Cylindrical feature** obtained from simplified PointNet : $\mathbb{R} \in C \times H \times W \times L$
  - C : feature dimension, H : radius, W : azimuth, L : height

- radius ϱ : distance to origin in xy axis
- azimuth θ : angle from x-axis to y-axis

Asymmetrical 3d convolution networks

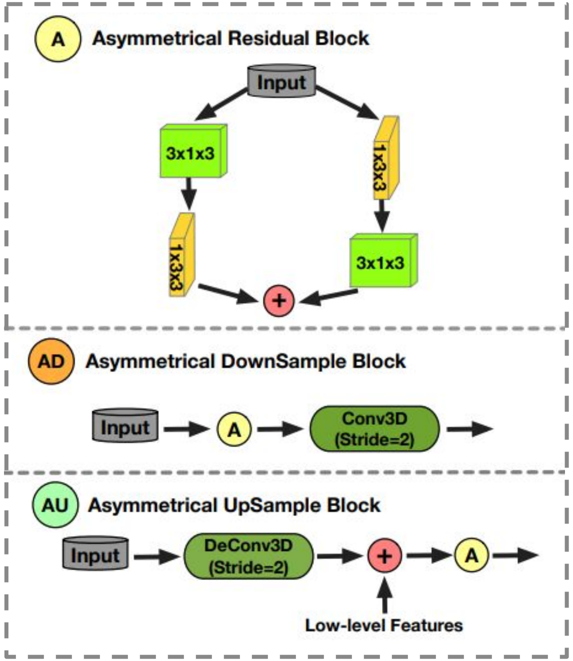




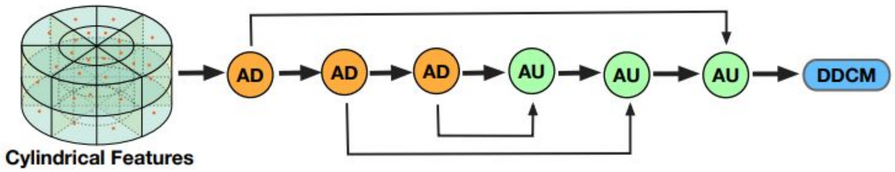
- Asymmetrical Residual Block [[Paper]()]

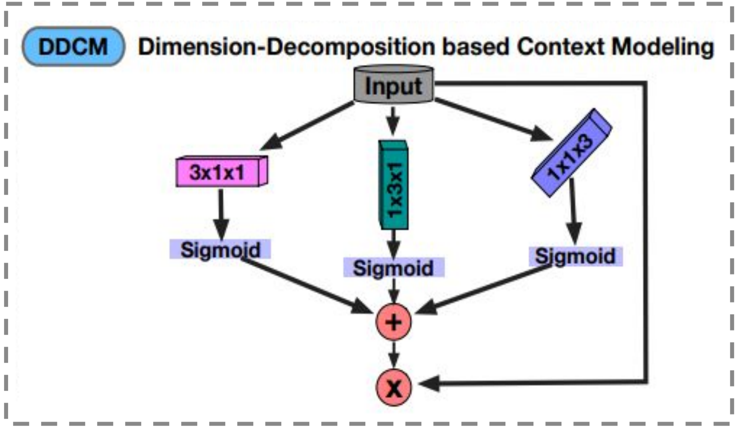
: strengthen the horizontal and vertical kernels, which matches the point distribution of object and makes the skeleton of the kernel powerful

- point distribution of object : driving scene point cloud carries out the specific object shape distribution like cubic object



Dimension-Decomposition based Context Modeling



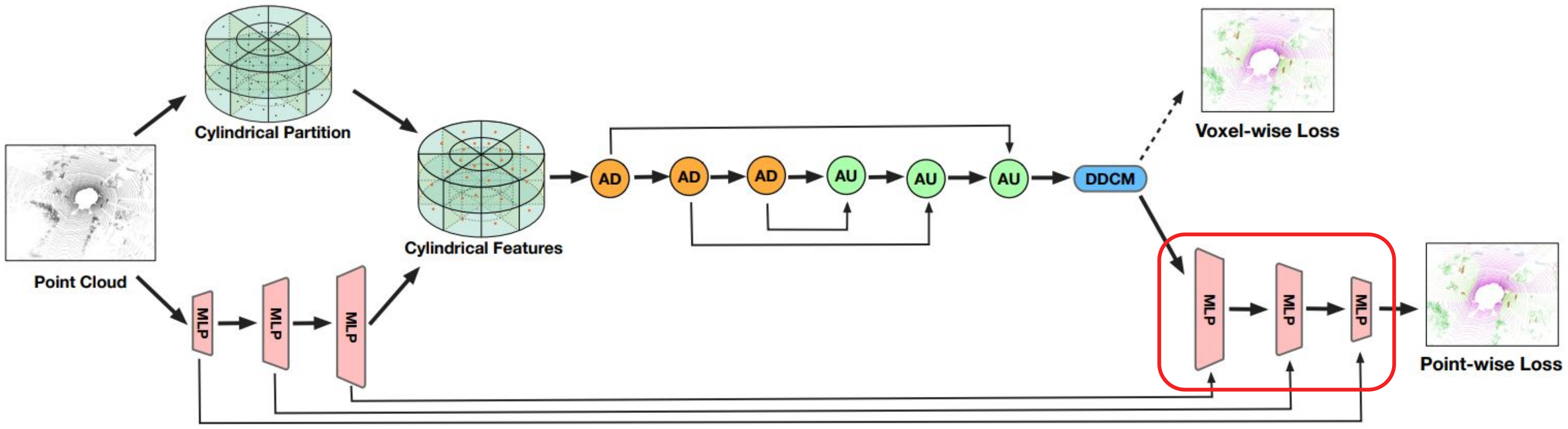


- Tensor decomposition theory [paper]

  : build the high-rank context as a combination of low-rank tensors

- three rank-1 kernel : obtain the low-rank features

Point-wise Refinement Module

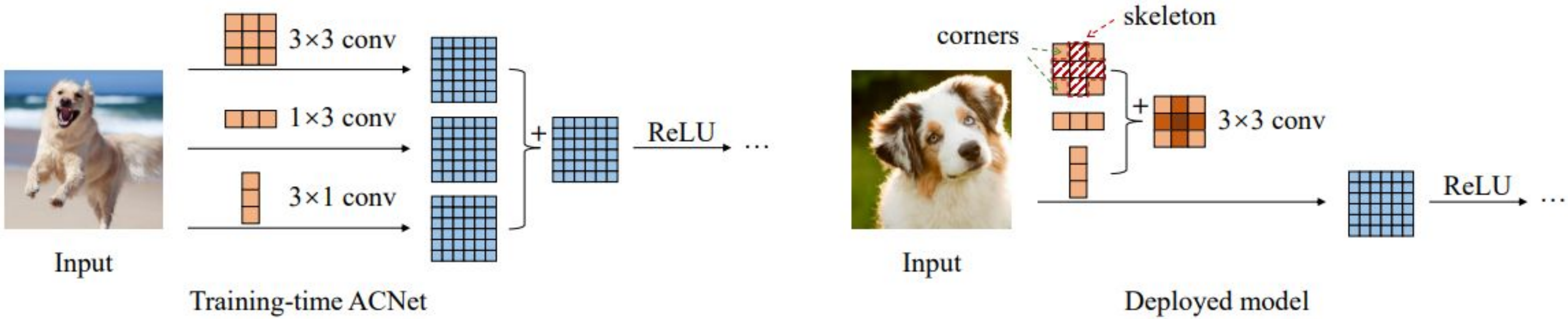


- Input : **both** point features before and after 3D convolution networks
- Project the voxel-wise features to the point-wise based on the point-voxel mapping table.

Paper Review

ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks

ICCV, 2019



ACNet

Goal

: Consider meaningful to improve the performance of CNN with **no extra inference-time computations**

- For an off-the-shelf architecture, it **replace** the standard square-kernel convolutional layers with **Asymmetric Convolution Blocks**

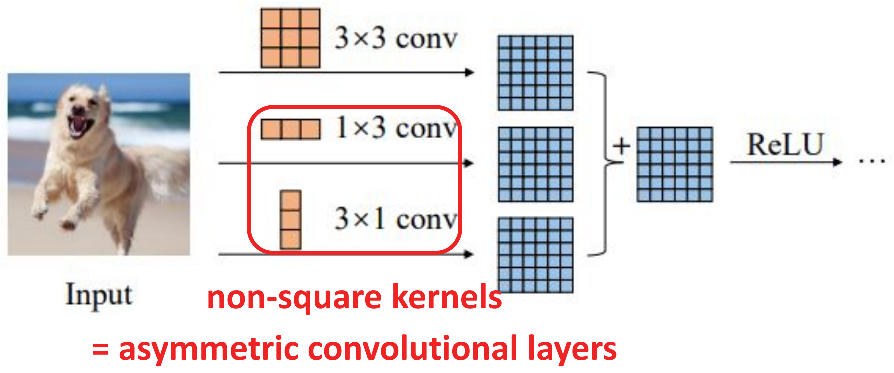
- After training, it equivalently convert the ACNet into the same original architecture, thus requiring no extra computations any more

Asymmetric Convolution Block(ACB)

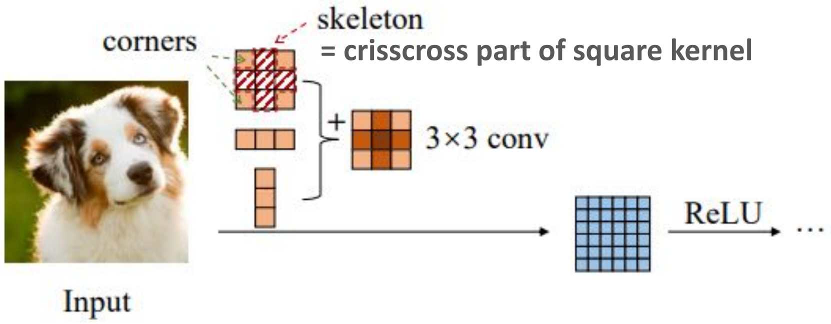
: 1D asymmetric convolutions(= horizontal and vertical kernels) to strengthen the square convolution kernels

Why it is work?

Square Convolution Kernel distributes its learned knowledge unequally, as the weights on the central crisscross position(=skeleton) are usually larger in magnitude



Training-time ACNet

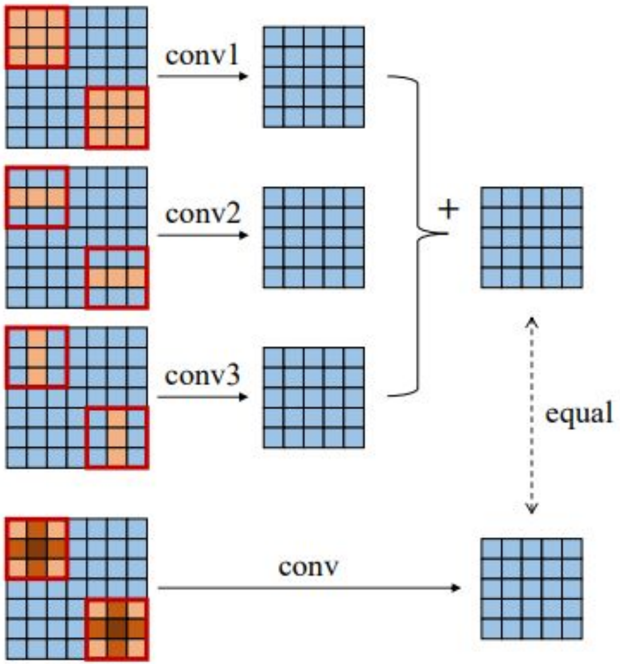


Deployed model

Why it is work?

- Property of convolution:

  if several 2D kernels with compatible sizes operate on the same input with the same stride to produce outputs of the same resolution, and their outputs are summed up, we can add up these kernels on the corresponding positions to obtain an equivalent kernel which will produce the same output.



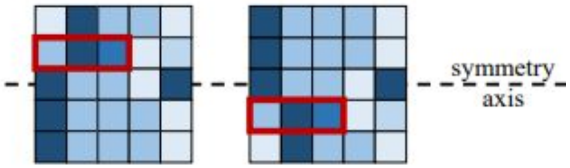
Ablation study

Horizontal Kernel

- Usually Argumentation : random left-right but no updown image flipping
- Case: Upside-down image is fed into the model

- Original 3 × 3 layers : produce meaningless results
- Horizontal kernel : produce the same outputs as on the original image at the axially symmetric locations

➡ Enhance the model’s robustness to rotational distortions by an observable margin.

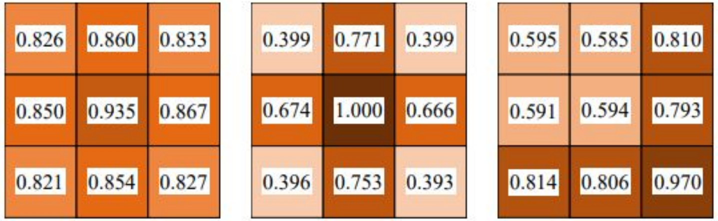


	Model	Horizontal kernel	Vertical kernel	Original input	Rotate 90°	Rotate 180°	Up-down flip
3	AlexNet		✓	57.10	29.65	32.86	33.02
2	AlexNet	✓		57.25	29.97	33.74	33.74
1	AlexNet	✓	✓	**57.44**	**30.49**	**33.98**	**33.82**
3	ResNet-18		✓	70.78	41.61	42.47	42.66
2	ResNet-18	✓		70.70	42.06	**43.22**	43.05
1	ResNet-18	✓	✓	**71.14**	**42.20**	42.89	**43.10**

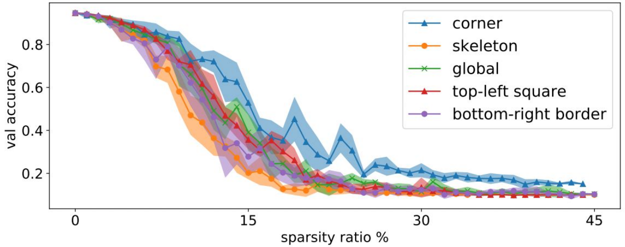
Table 4: Top-1 accuracy of the ACNets with different design configurations and rotational distortions on ImageNet.

Ablation study

Skeleton vs Border



(a) Normal. (b) ACNet, skeleton. (c) ACNet, border.



(c) ACNet with the asymmetric kernels added to the border.

: Increase the magnitude of the borders but cannot diminish the importance of the other parts.

- Pruning experiments

- Pruning the corners : still delivers the best accuracy
- Pruning the borders : no better results than the top-left 2 × 2 squares

Result

Table 1: Top-1 accuracy of ACNets and the normally trained baselines on CIFAR-10.

Model	Base Top-1	ACNet Top-1	Top-1 ↑
Cifar-quick	83.13	84.24	1.11
VGG	94.12	94.47	0.35
ResNet-56	94.31	95.09	0.78
WRN-16-8	95.56	96.15	0.59
DenseNet-40	94.29	94.84	0.55

Table 2: Top-1 accuracy of ACNets and the normally trained baselines on CIFAR-100.

Model	Base Top-1	ACNet Top-1	Top-1 ↑
Cifar-quick	53.22	54.30	1.08
VGG	74.56	75.20	0.64
ResNet-56	73.58	74.04	0.46
WRN-16-8	78.65	79.44	0.79
DenseNet-40	73.14	73.41	0.27

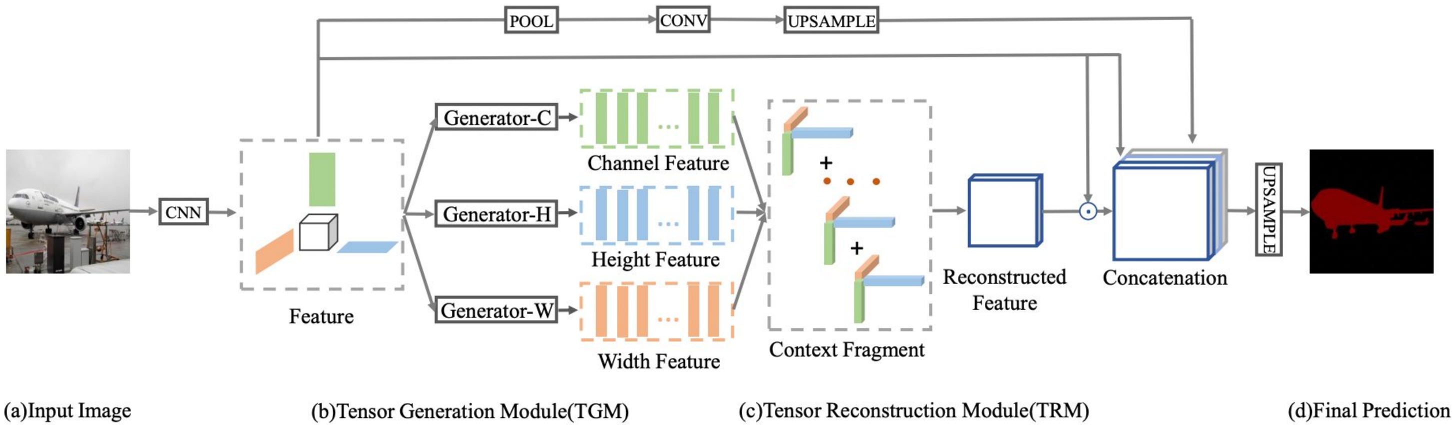
Table 3: Accuracy of the ACNet counterparts of AlexNet, ResNets, DenseNet-121 and the baselines on ImageNet.

Model	Base Top-1	ACNet Top-1	Top-1 ↑	Base Top-5	ACNet Top-5	Top-5 ↑
AlexNet	55.92	**57.44**	1.52	79.53	**80.73**	1.20
ResNet-18	70.36	**71.14**	0.78	89.61	**89.96**	0.35
DenseNet-121	75.15	**75.82**	0.67	92.45	**92.77**	0.32

Paper Review

Tensor Low-Rank Reconstruction for Semantic Segmentation

ECCV, 2020



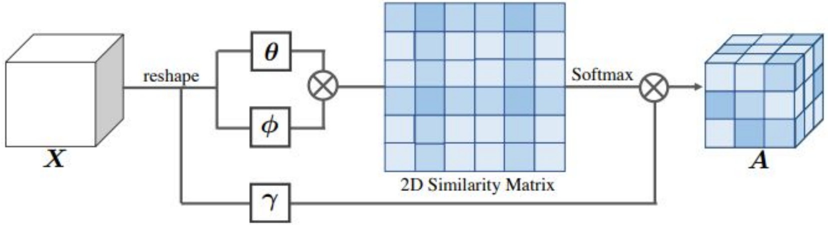
Goal

- Non-local self-attention

  : proved to be effective for context information collection

- Context = spatial-wise attention + channel-wise attentions (= 3D information)

1. 2D similarity matrix : channel-wise attention missing because of Space compression
2. Directly use context without compression : high-rank difficulty of context



- Goal

  : 3D context representations, which not only avoids the space compression but also tackles the high-rank difficulty.

Canonical-Polyadic Decomposition

How can we obtain the most similar M to the input data X?

: High-rank tensor can be expressed as a combination of rank-1 tensors

The Tensor generated by one Vector multiplication cannot be completely identical to the left Tensor due to the difference in information amount. Therefore, the unresolved error is solved by adding the remaining Vector multiplication.