DECOUPLED DYNAMIC FILTER NETWORKS

Jingkai Zhou[12] Varun Jampani[3] Zhixiong Pi[24] Qiong Liu[1] Ming-Hsuan Yang[235]

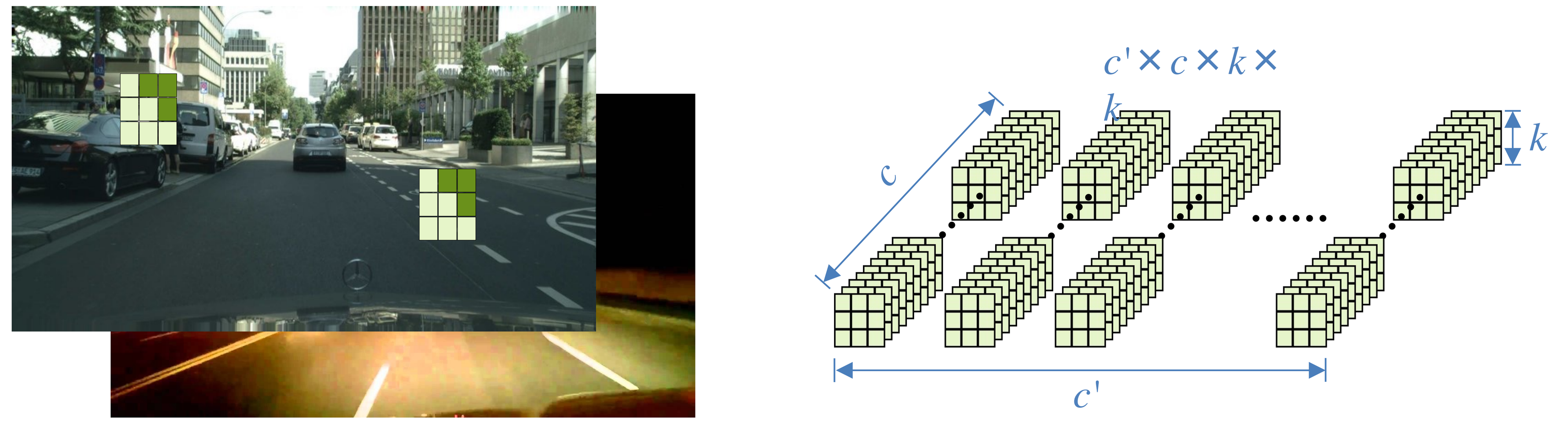
1 South China University of Technology 2 University of California at Merced 3 Google Research 4 Huazhong University of Science and Technology 5 Yonsei University

CVPR VIRTUAL JUNE 19-25

DDF is a light-weight, high-performing content-adaptive convolution layer that can readily replace standard convolution layers in CNNs.

1. Introduction

Problem: Convolution has two short-comings.

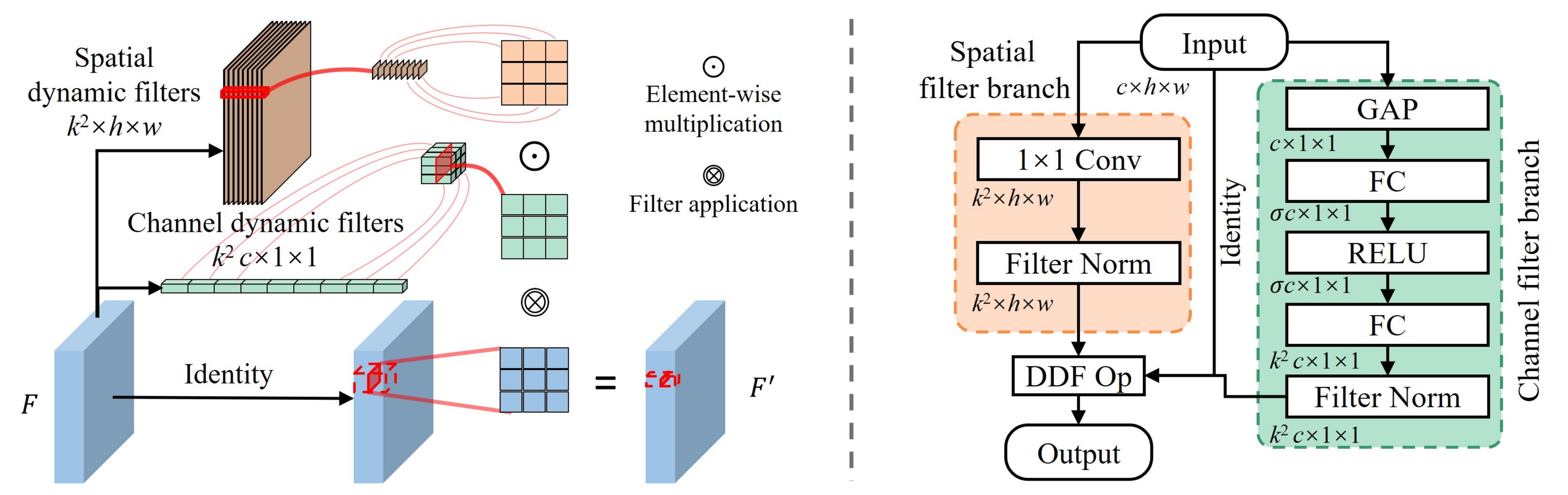


Content-agnostic　　Computation heavy

- Dynamic filters tackle the first issue while further increasing the computational costs.
- Grouped/depthwise convolution reduce the computational costs, which usually result in a drop of performance.

Goal: Design a filtering operation that is content-adaptive while also being lighter-weight than a standard convolution.

2. DDF Module

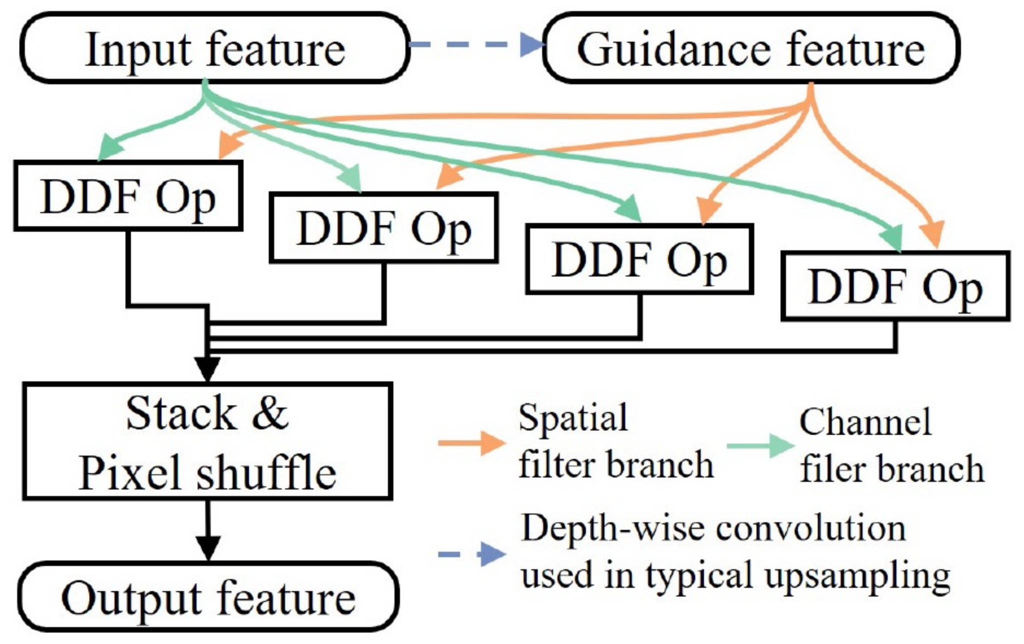


Key idea: Decouple dynamic filters into spatial and channel filters.

- First, predict spatial/channel filters individually via two side-branches.
- Then, combine spatial/channel filters at each pixel and channel.
- At last, apply the combined filter on the corresponding location of the input feature.

Formulation: $F'_{(r,i)} = \sum_{j \in \Omega(i)} D_i^{sp}[\mathbf{p}_i - \mathbf{p}_j] D_r^{ch}[\mathbf{p}_i - \mathbf{p}_j] F_{(r,j)}$

3. DDF-Up Module



Propose a unified DDF-Up module for typical/joint upsampling task.

- 4 branches for scale factor 2. Stacking multiple DDF-Up for larger scale factor.
- Apply spatial/channel branch on guided/input feature, respectively
- For typical upsampling, a depth-wise convolution is used to generate guided feature

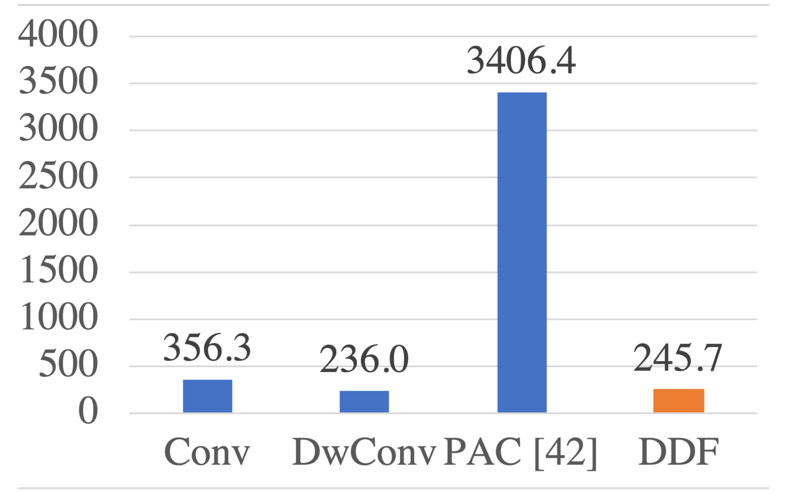
4. Computational Complexity

Table 1. Theoretical computational cost.

Filter	Conv	DwConv	DyFilter	DDF
Params	c^2k^2	ck^2	c^3k^2	$ck^2 + \sigma c^2(1+k^2)$
Time	$O(nc^2k^2)$	$O(nck^2)$	$O(nc^3k^2)$	$O(nck^2 + c^2k^2)$
Space	–	–	$O(nc^2k^2)$	$O((n+c)k^2)$

Table 2. Latency on different resolutions.

Resolution	Conv	DwConv	DDF	DDF Op
7×7	0.21 ms	0.05 ms	0.93 ms	0.12 ms
14×14	0.40 ms	0.09 ms	0.96 ms	0.15 ms
28×28	2.31 ms	0.22 ms	1.29 ms	0.48 ms
56×56	4.09 ms	0.79 ms	2.60 ms	1.80 ms
112×112	16.04 ms	3.08 ms	9.07 ms	7.30 ms
224×224	82.57 ms	11.97 ms	37.11 ms	28.62 ms



Runtime memory (M).

Table 3. Ablation study on branches.

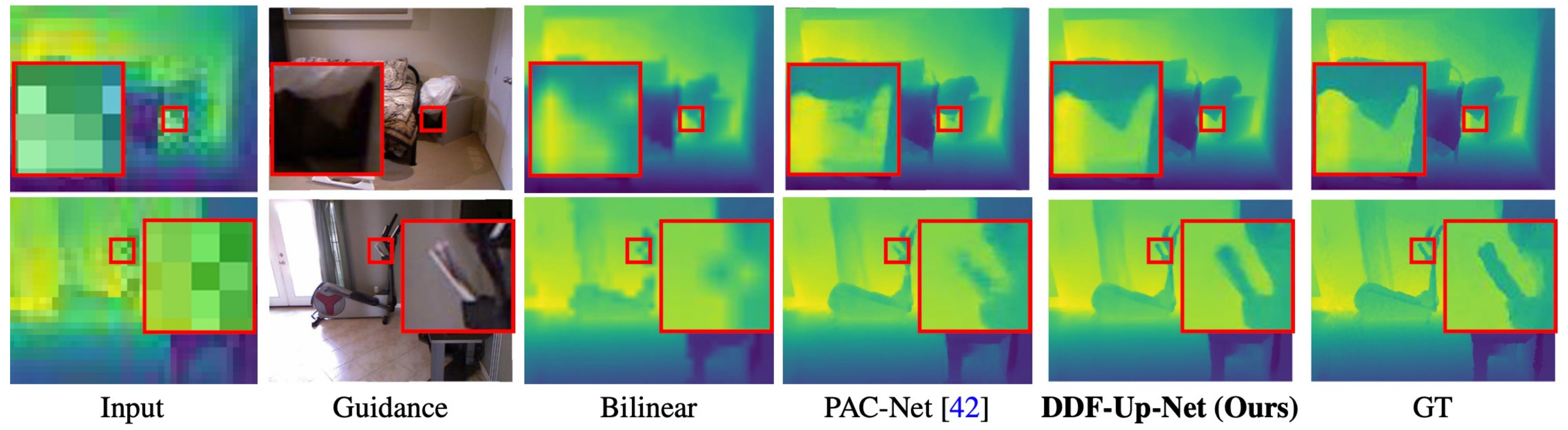
Spatial	Channel	Top-1 / Top-5 Acc.
Base Model		77.2 / 93.5
✓		74.4 / 92.0
	✓	78.7 / 94.2
✓	✓	**79.1 / 94.5**

Table 4. Comparisons between filters.

Arch	Conv Type	Params	FLOPs	Top-1 Acc
R18	*Base Model* [15]	11.7M	1.8B	69.6
	Adaptive [57]	11.1M	–	70.2
	DyNet [59]	16.6M	0.6B	69.0
	DDF	**7.7M**	**0.4B**	**70.6**
R50	*Base Model* [15]	25.6M	4.1B	77.2
	DyNet [59]	–	**1.1B**	76.3
	CondConv [55]	104.8M	4.2B	78.6
	DwCondConv [55]	14.5M	2.3B	78.3
	DwWeightNet [34]	**14.4M**	2.3B	78.0
	DDF	16.8M	2.3B	**79.1**

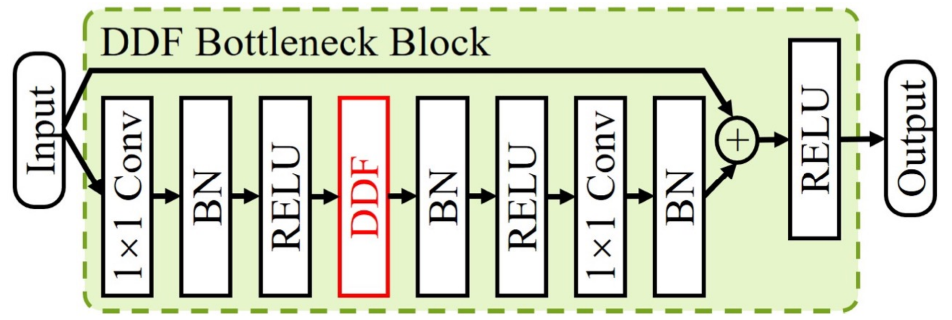
Table 5. Comparison with variants of ResNets on the ImageNet 1K.

Method	Params	FLOPs	Top-1 Acc
ResNet50 (base) [15]	25.6M	4.1B	76.0 (77.2)
SE-ResNet50 [18]	28.1M	4.1B	77.6 (77.8)
BAM-ResNet50 [36]	25.9M	4.2B	76.0
CBAM-ResNet50 [50]	28.1M	4.1B	77.3
AA-ResNet50 [2]	25.8M	4.2B	77.7
ResNeXt50 (32×4d) [53]	25.0M	4.3B	77.8 (78.2)
Res2Net50 (14w-8s) [12]	25.7M	4.2B	78.0
DDF-ResNet50	**16.8M**	**2.3B**	**79.1**
ResNet101 (base) [15]	44.5M	7.8B	77.6 (78.9)
SE-ResNet101 [18]	49.3M	7.8B	78.3 (79.3)
BAM-ResNet101 [50]	44.9M	7.9B	77.6
CBAM-ResNet101 [50]	49.3M	7.8B	78.5
AA-ResNet101 [2]	45.4M	8.1B	78.7
ResNeXt101 (32×4d) [53]	44.2M	8.0B	78.8 (79.5)
Res2Net101 (26w-4s) [12]	45.2M	8.1B	79.2
DDF-ResNet101	**28.1M**	**4.1B**	**80.2**

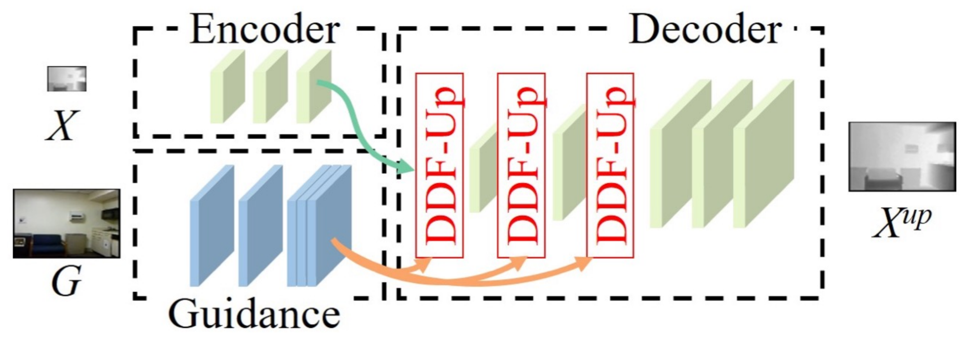


We visualize 16 times joint depth upsampling results, where we can see that DDF-Up-Net recovers more details compared to PAC-Net and other techniques.

5. Experiments



DDF bottleneck block for classification



DDF-Up-Net for joint depth upsampling

6. Conclusion

Propose the DDF and DDF-Up modules which have the following favorable properties:

- Content adaptive.
- Fast with small memory footprint.
- Consistent improvements on different tasks.
- Can be used as a basic building block.

Project Page