FPGA-based Real-Time Super-Resolution System for Ultra High Definition Videos

Zhuolun He, Hanxian Huang, Ming Jiang, Yuanchao Bai, and Guojie Luo

Peking University

FCCM 2018

Ultra High Definition (UHD) Technology

UHD Television

UHD Projector

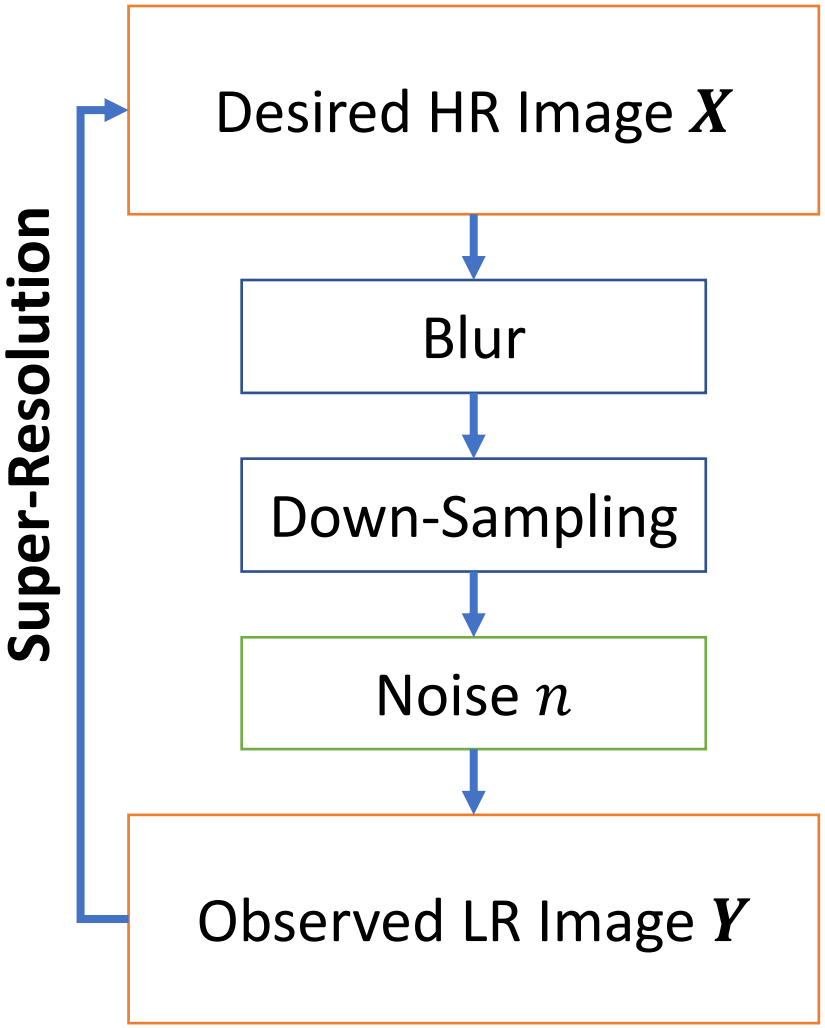
Content?

- Limited Creators
- High network bandwidth cost
- Huge storage cost

UHD Phone

UHD Camera

High-Resolution <---> Low-Resolution



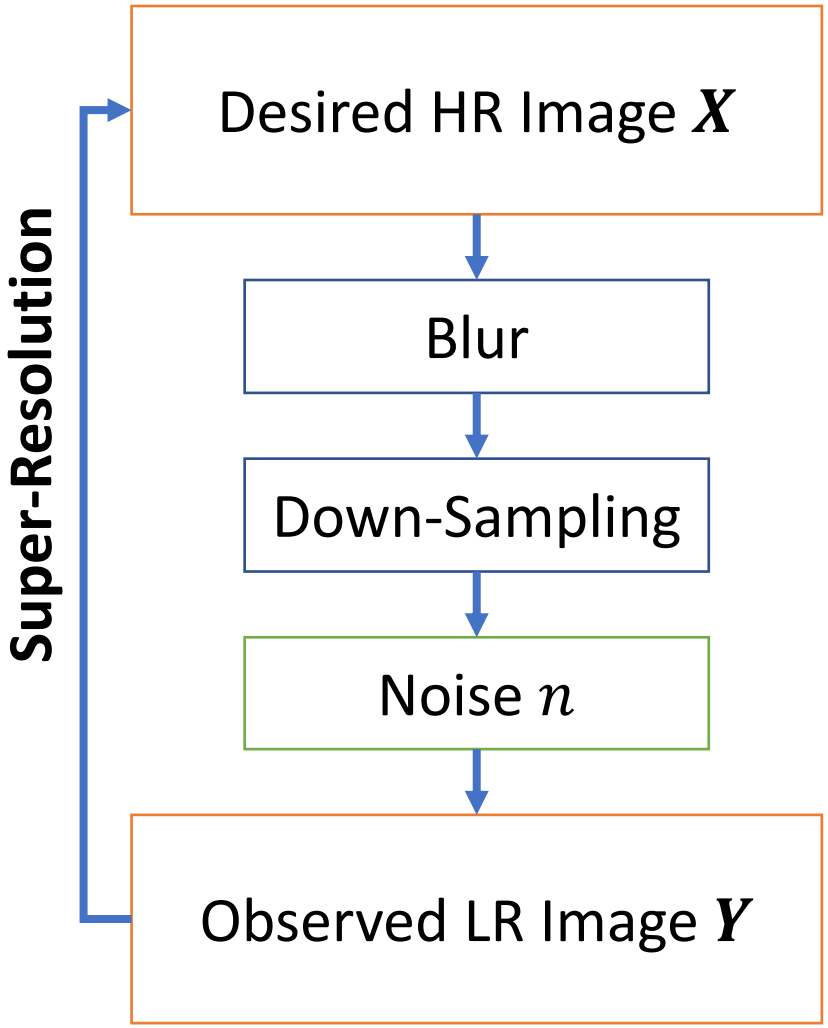
Spectrum of Super Resolution Methods

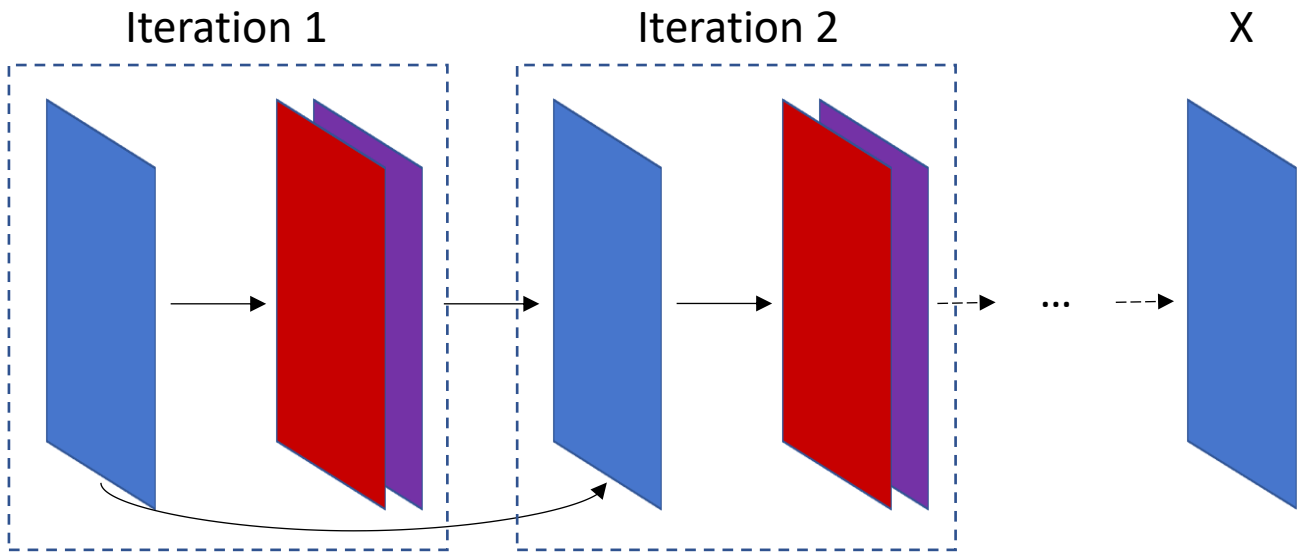
Simple

Complicated

Interpolation	Model-based	Example-based
• Fast • Easy to implement • Blurry results	• Interpretable • High complexity • Assumed known blur kernel/noise	• State-of-the-art quality • High complexity • Training data needed

Model-based Method is also Compute-Intensive

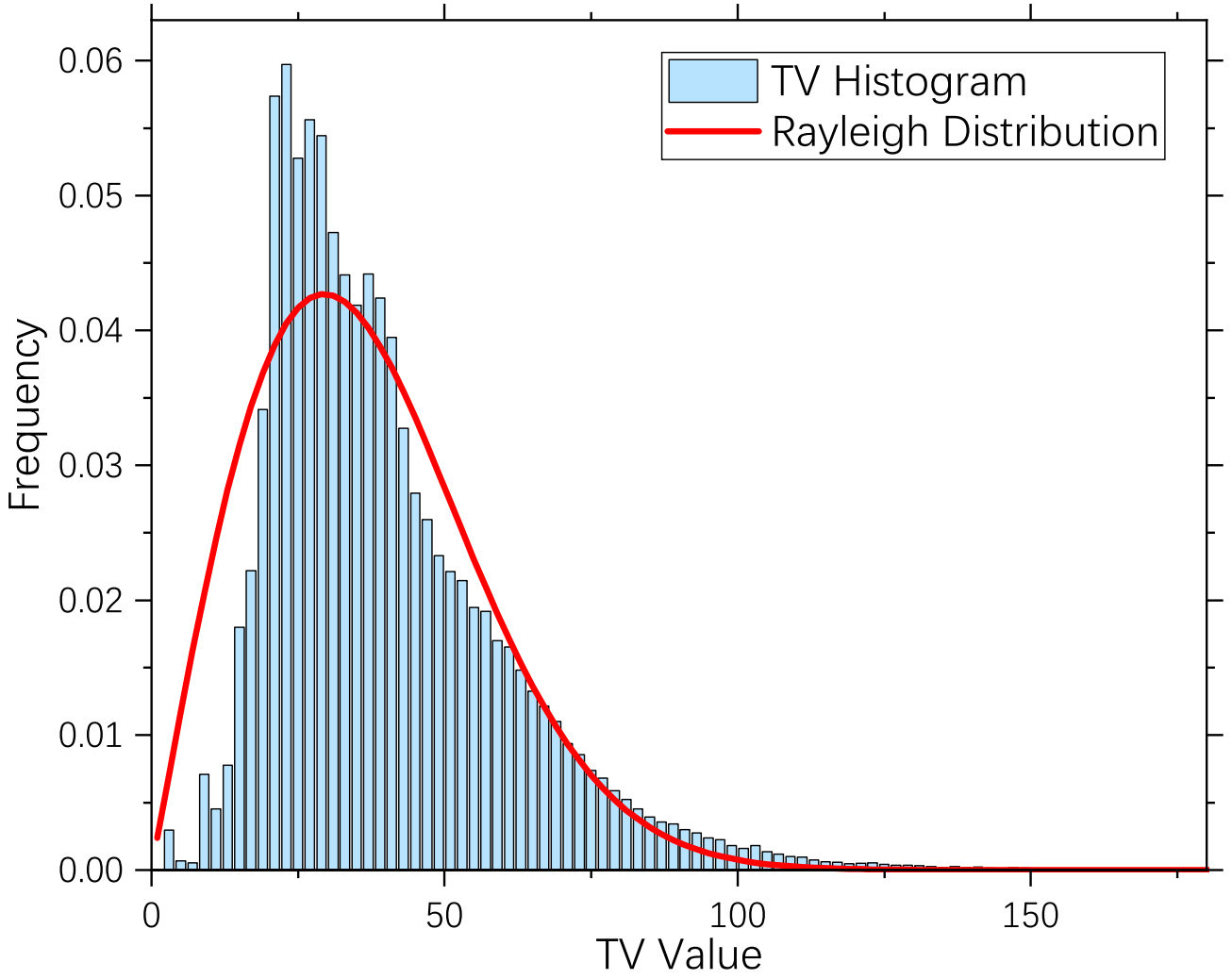




Model-based methods may not be needed

- The computation also has a layered structure
- We can use a neural network to approximate

Total Variation Distribution



Fact:

Blocks contain DIFFERENT amount of information (NOT equally important)

Insight:

Use DIFFERENT upscaling methods for different blocks

A Hybrid Algorithm

INPUT: LR Image $\boldsymbol{Y}$

1. Crop $\boldsymbol{Y}$ into sub-images $\{\boldsymbol{y}\}$

2.1. $\boldsymbol{x}$ <- Upscale($\boldsymbol{y}$) IF $\boldsymbol{M}(\boldsymbol{x}) > \boldsymbol{T}$

2.2. ELSE $\boldsymbol{x}$ <- CheapUpscale($\boldsymbol{y}$)

3. Mosaic $\boldsymbol{X}$ with $\{\boldsymbol{x}\}$

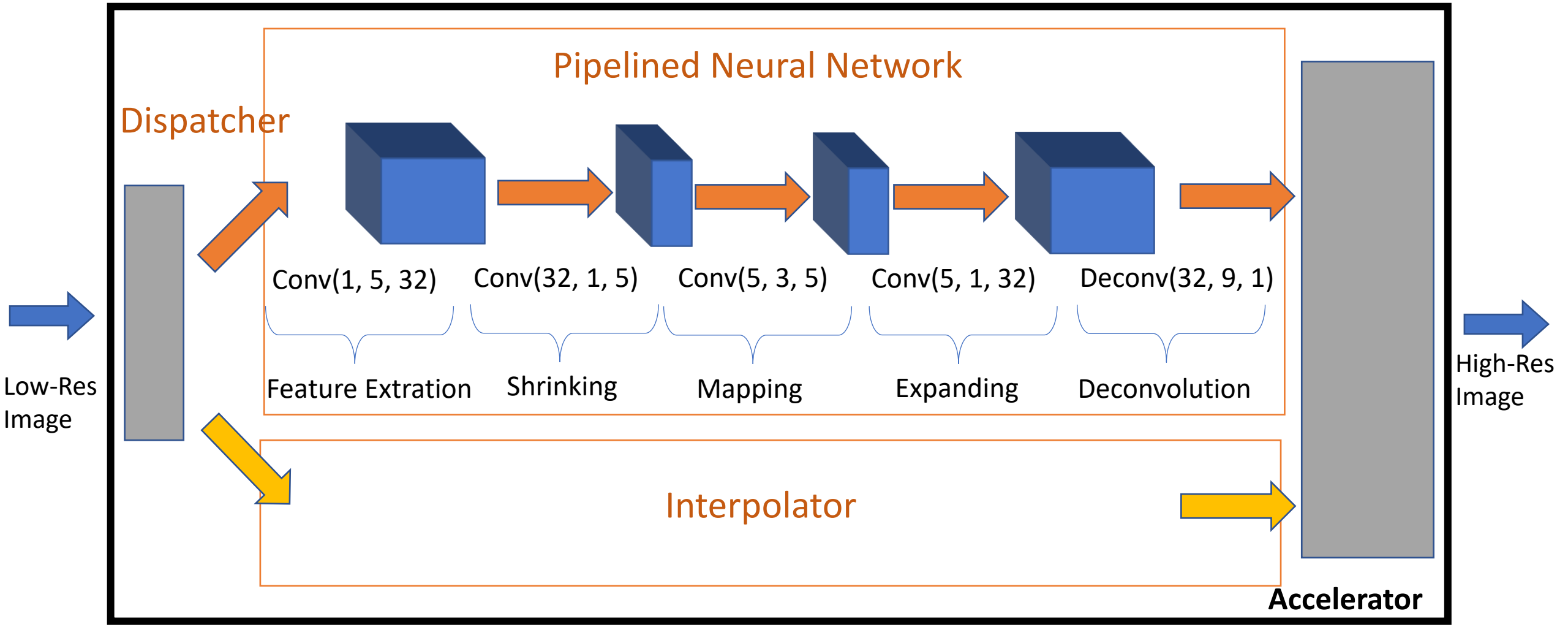
OUTPUT: HR Image $\boldsymbol{X}$

M: Total Variation (TV)

Upscale: FSRCNN-s

CheapUpscale: Intepolation

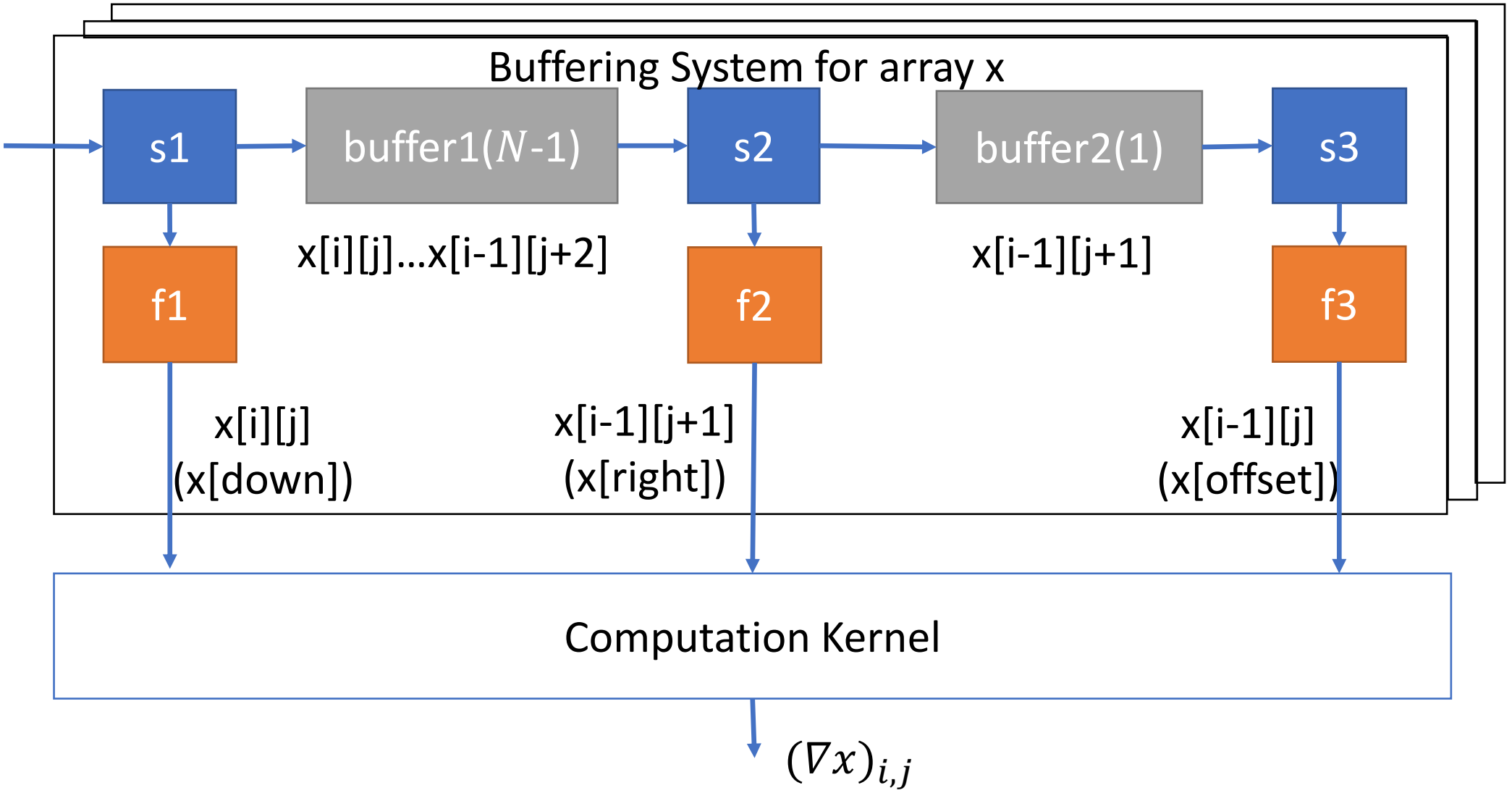
Overall System



Stencil Access of TV Computation

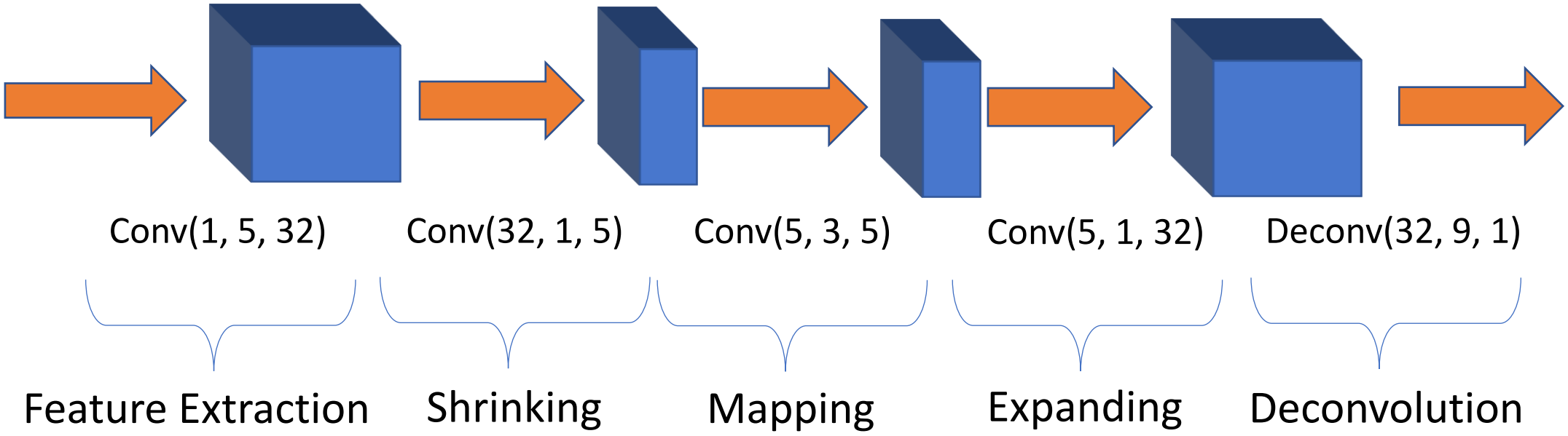
$$(\nabla x)_{\text{offset}} = abs(x[\text{right}] - x[\text{offset}]) + abs(x[\text{down}] - x[\text{offset}])$$

Micro-architecture for Stencil Computation

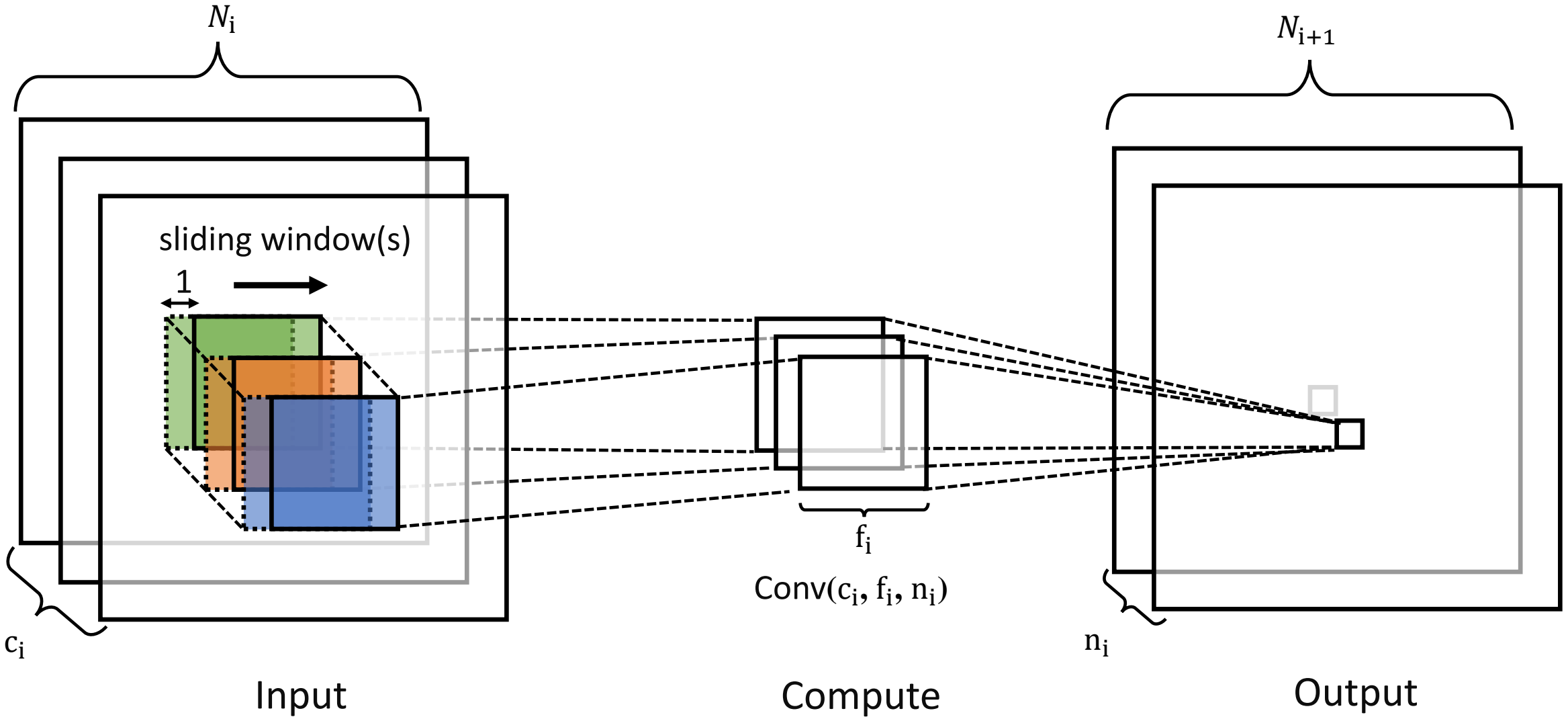


Convolutional Neural Network

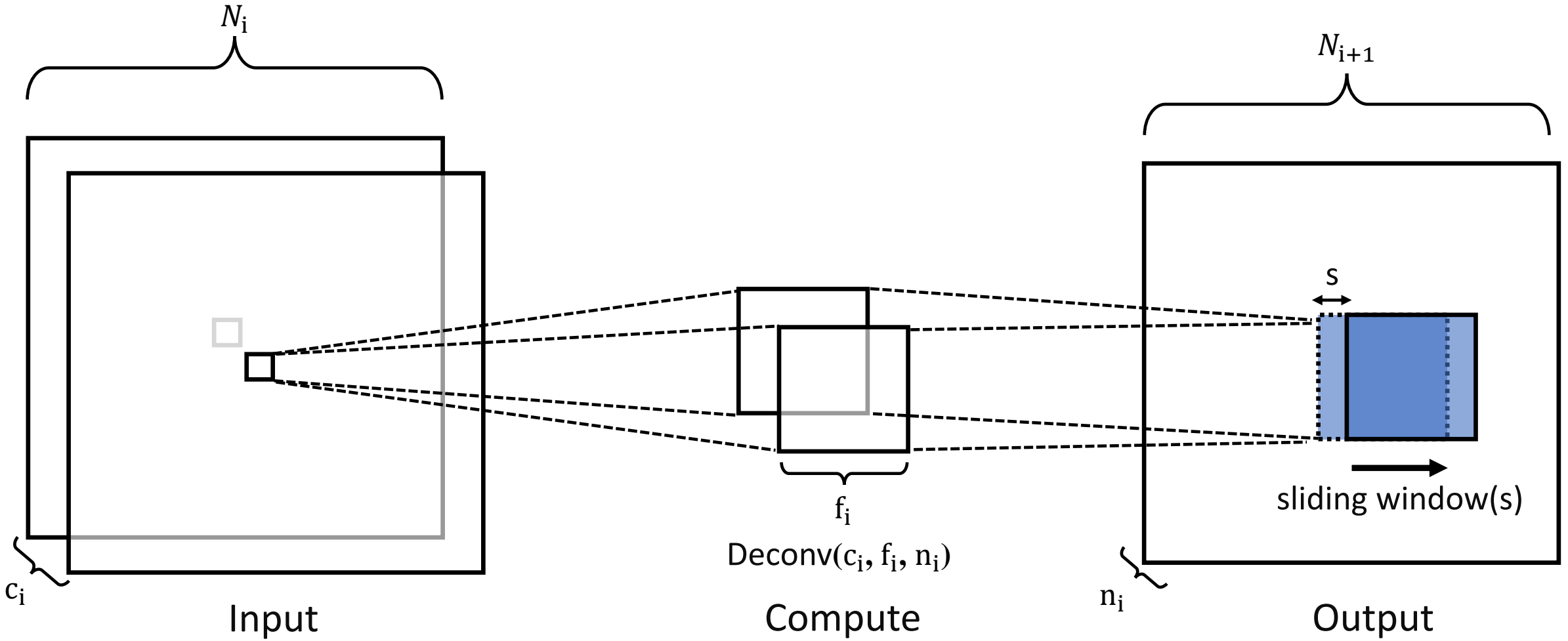
Pipelined Neural Network



Convolution



Deconvolution



Pipeline Balancing

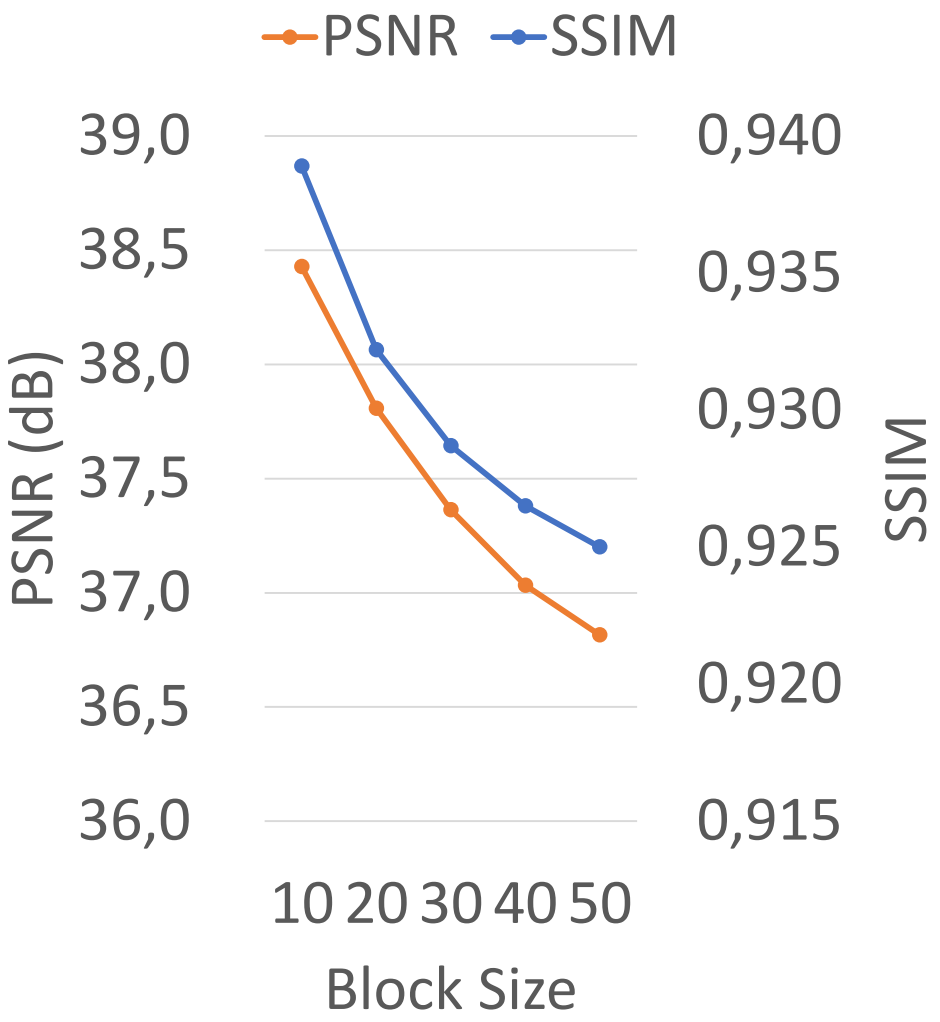
Layer	c_i	f_i	n_i	N_i	#Mult.	Ideal #DSP	Ideal II	Alloc. #DSP	Alloc. II
Extraction	1	5	32	36	819200	201	4076	200	4096
Shrinking	32	1	5	32	163840	40	4096	32	4096
Mapping	5	3	5	32	202500	50	4050	45	4500
Expanding	5	1	32	30	144000	35	4115	32	4500
Deconvolution	32	9	1	30	2332800	573	4072	519	4500
Overall	-	-	-	-	3662340	899	4115	828	4500
Available (ZC706)	-	-	-	-	-	900	-	900	-

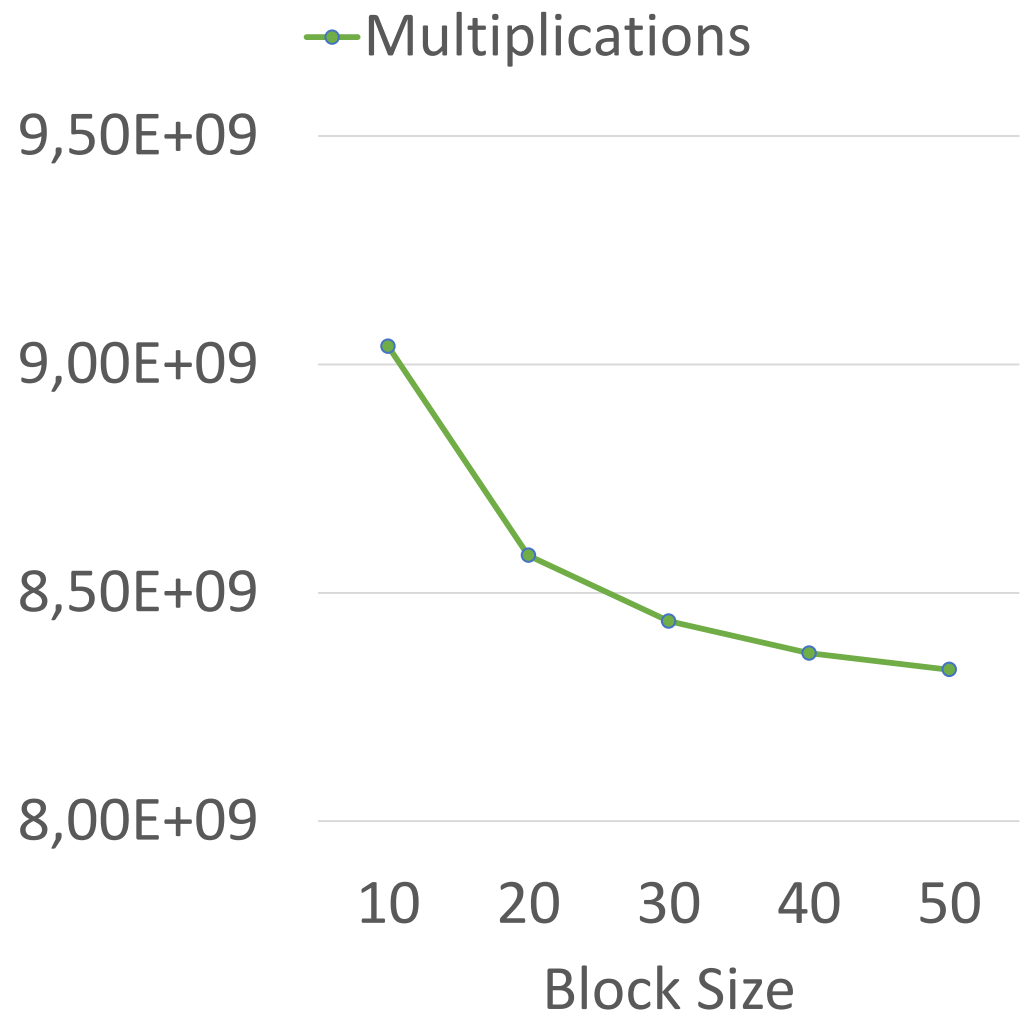
Sub-image Size

- Padding
  - $N_i \equiv k + \sum_i^{\#Conv}(f_i - 1)$

- If sub-image size too small
  - Large border-to-block ratio
  - Limited by memory bandwidth
- If sub-image size too large
  - Large feature maps
  - Limited by on-chip BRAM capacity

Sub-image Size vs. Performance vs. #mult.





Overall Comparisons

- Compared six configurations

No.	Preprocessing	Upscaling	#Mult.	PSNR(dB)	SSIM
1	None	Interpolation	6.6*10^7	35.51	0.9138
2	None	Neural Network	8.2*10^9	38.55	0.9421
3	Blocking	Interpolation	6.6*10^7	35.51	0.9138
4	Blocking	Neural Network	8.4*10^9	38.55	0.9420
5	Blocking	Mixed-Random	2.2*10^9	36.10	0.9211
6	Blocking	Mixed-TV	2.2*10^9	37.36	0.9287

>100x

+3.04dB

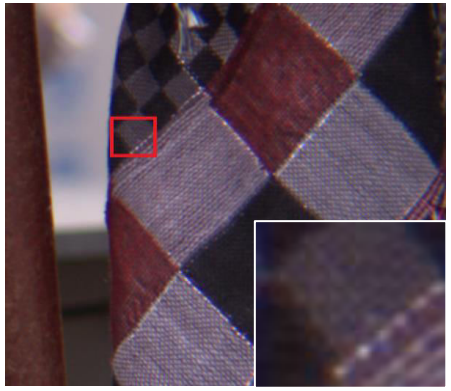
No Performance Loss

+1.26dB

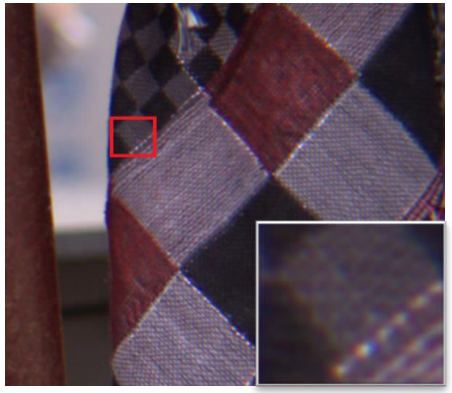
-75%

-1.19dB

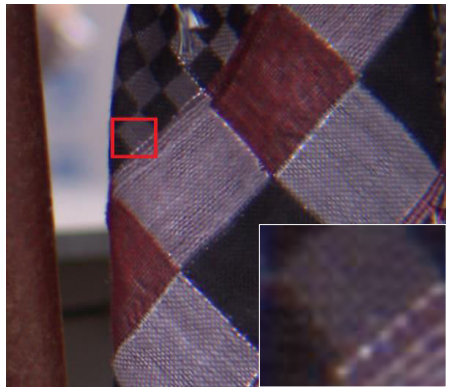
Example Outputs



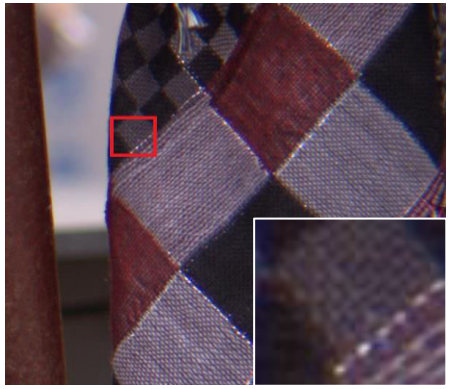
Configuration 1
None/Interpolation



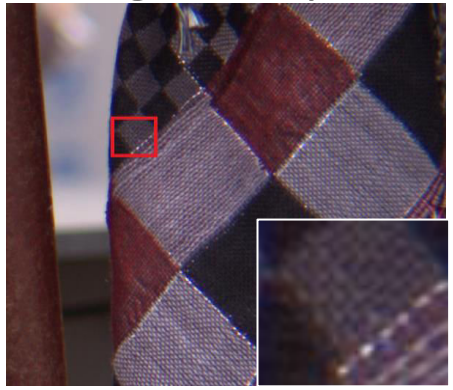
Configuration 3
Blocking/Interpolation



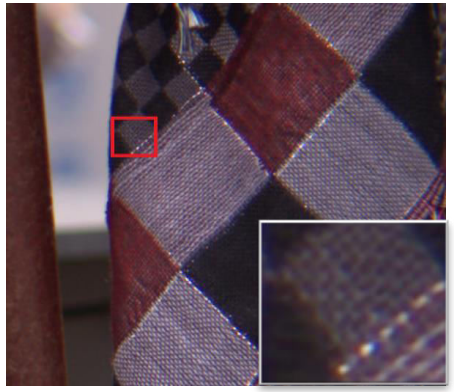
Configuration 5
Blocking/Mixed-Random



Configuration 2
None/Neural Network



Configuration 4
Blocking/Neural Network



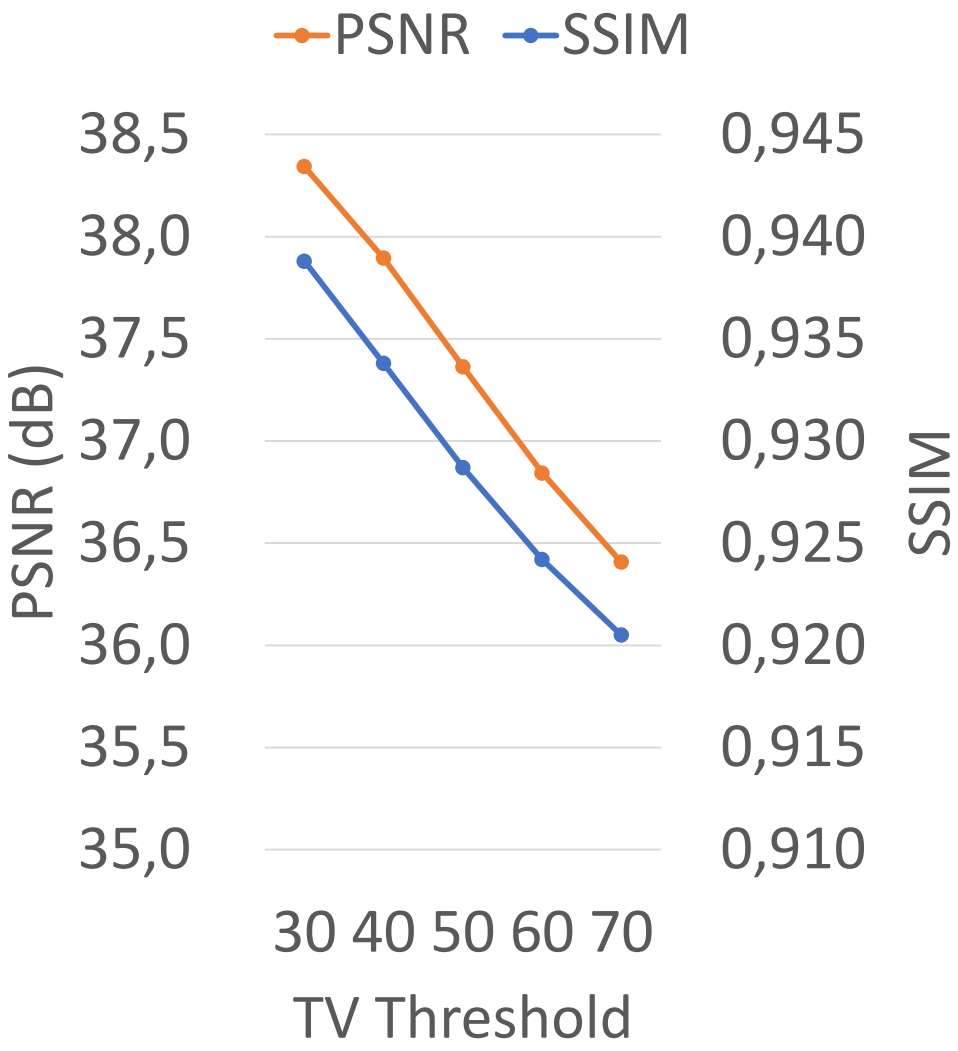
Configuration 6
Blocking/Mixed-TV

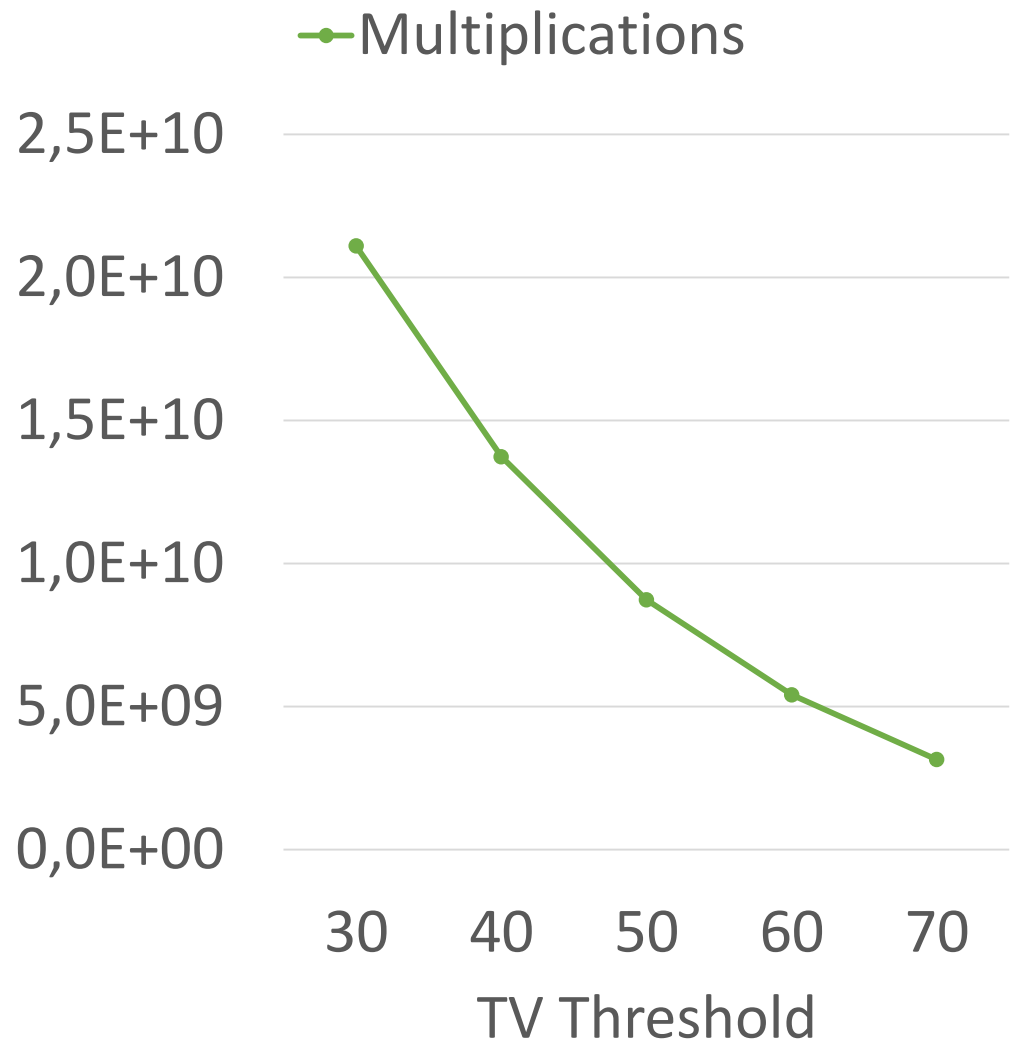
Summary Flow

- Crop each frame into blocks
  - Suitable for low-level (pixel-level) tasks
  - GOOD: on-chip buffer friendly
  - BAD: Computation overheads
- Dispatch blocks according to TV value
  - Micro-architecture for buffering system
- Fully-pipelined CNN for upscaling
  - Sliding window for convolution/deconvolulution
  - Pipeline balancing
- Performance
  - Full-HD (1920x1080) -> Ultra-HD (3940x2160): 31.7fps

Thank you!

TV Threshold vs. Performance vs. #mult.





Resource Utilizations

Component	BRAM	DSP	FF	LUT
Dispatcher	1	2	618	1138
Neural Network	178	844	63149	98439
Interpolator	0	10	1414	3076
Total	327	858	66261	103714
Available	1090	900	437200	218600
Utilization (%)	**30**	**95**	**15**	**47**