Poster Overview

SOGANG UNIVERSITY
Machine Learning Laboratory

N-Gram in Swin Transformers for Efficient Lightweight Image Super-Resolution

Haram Choi[1], Jeongmin Lee[2], and Jihoon Yang[1]

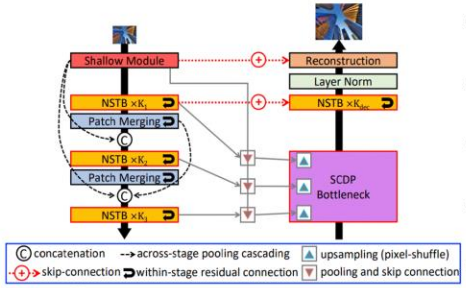
[1]Dept. of Computer Science & Engineering, Sogang University [2]LG Innotek, Seoul, Republic of Korea

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

❖ Motivation

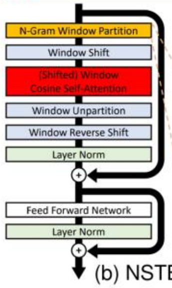
- Receptive field of Swin Transformer is too limited.
- SR requires the information of locally adjacent regions to recover degraded pixels.
- The popular Vision Transformer family originates from Transformer of NLP domain.
- N-Gram language model can consider longer spans in text analysis.
- Inspired by such advantage of N-Gram from NLP field, we propose N-Gram context in low-level vision task.

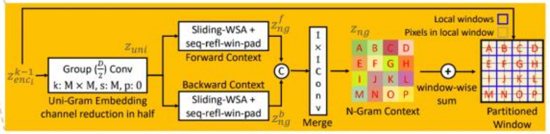
❖ Overall Architecture of NGswin (N-Gram Swin Transformer)



(a) NGswin

- *NSTB* (b) contains our proposed *N-Gram Window Partitioning* method.
- *Uni-Gram Embedding reduces feature maps* for efficiency.
- For *N-Gram context, sliding-WSA* (window self-attention) is used.
- *SCDP* enables NGswin to maintain an *efficient hierarchical structure.*
- Our method can *overcome the limited receptive fields* of Swin Transformers.

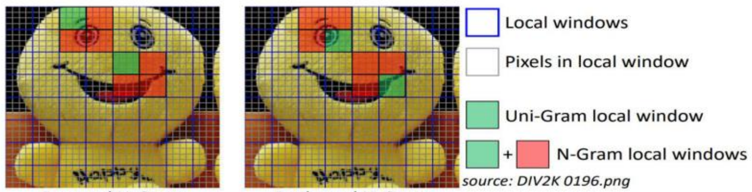




(c) N-Gram Window Partition

(b) NSTB (N-Gram Swin Transformer Block)

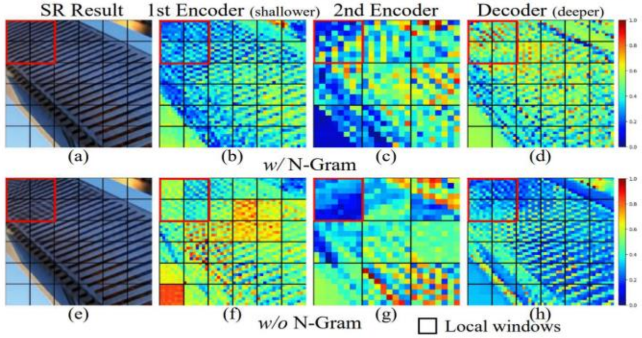
❖ Definition of N-Gram in an Image



Forward N-Gram Backward N-Gram

source: DIV2K 0196.png

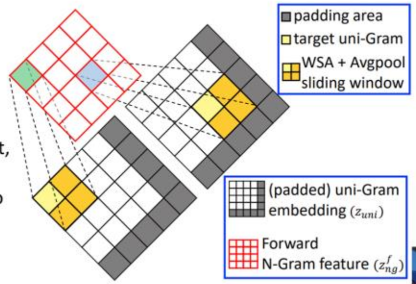
❖ Problem Verification (limited receptive field of Swin Transformer)



- **(f), (g)**: Self-attention in shallower layer is *limited within a local window.*
- **(h)**: The *patterns* in the red box and its neighbors *differ.*
- **(e)**: The *obvious distortion* in the red box is led by this issue.
- **(a), (b), (c), (d)**: However, our *N-Gram context can overcome this issue.*

❖ Sliding-WSA

- Example for *forward bi-Gram* context.
- For *backward bi-Gram* context, only the *direction of padding* is changed from lower-right to *upper-left*.

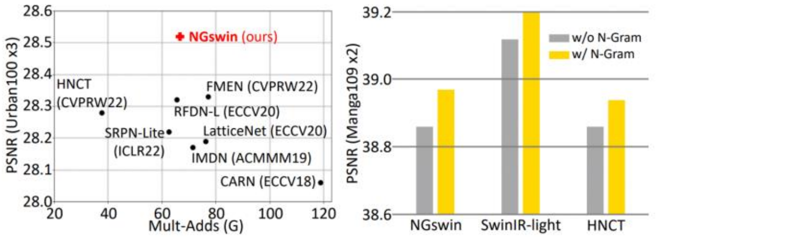


❖ SCDP Bottleneck (pixel-Shuffle / Concat / Depth-wise / Point-wise)

(b) Impacts of extra stages and SCDP bottleneck. PSNR / SSIM.

Stages	SCDP	Scale	Mult-Adds	#Params	Urban100	Manga109
extra	*w/o*	×2	87.98G	997K	32.28 / 0.9298	38.72 / 0.9773
default	*w/o*		138.88G	992K	32.48 / 0.9321	38.92 / 0.9776
default	***w/***		**140.41G**	**998K**	**32.53 / 0.9324**	**38.97 / 0.9777**
extra	*w/o*	×3	42.10G	1,006K	28.33 / 0.8562	33.67 / 0.9453
default	*w/o*		65.85G	1,001K	28.47 / 0.8596	33.81 / 0.9464
default	***w/***		**66.56G**	**1,007K**	**28.52 / 0.8603**	**33.89 / 0.9470**
extra	*w/o*	×4	23.33G	1,018K	26.22 / 0.7900	30.46 / 0.9090
default	*w/o*		36.06G	1,013K	26.38 / 0.7954	30.71 / 0.9121
default	***w/***		**36.44G**	**1,019K**	**26.45 / 0.7963**	**30.80 / 0.9128**

❖ Two Tracks of Paper

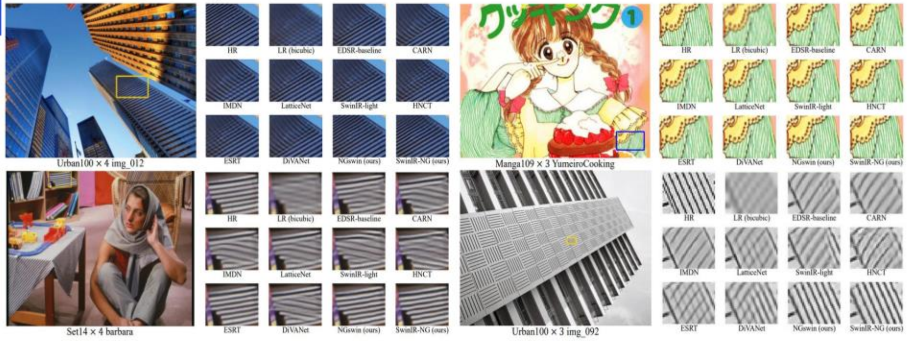


- **(left)**: NGswin outperformed previous state-of-the-art efficient Super-Resolution (SR) networks with a more efficient architecture.
- **(right)**: The proposed N-Gram context enhances other Swin Transformer-based networks.

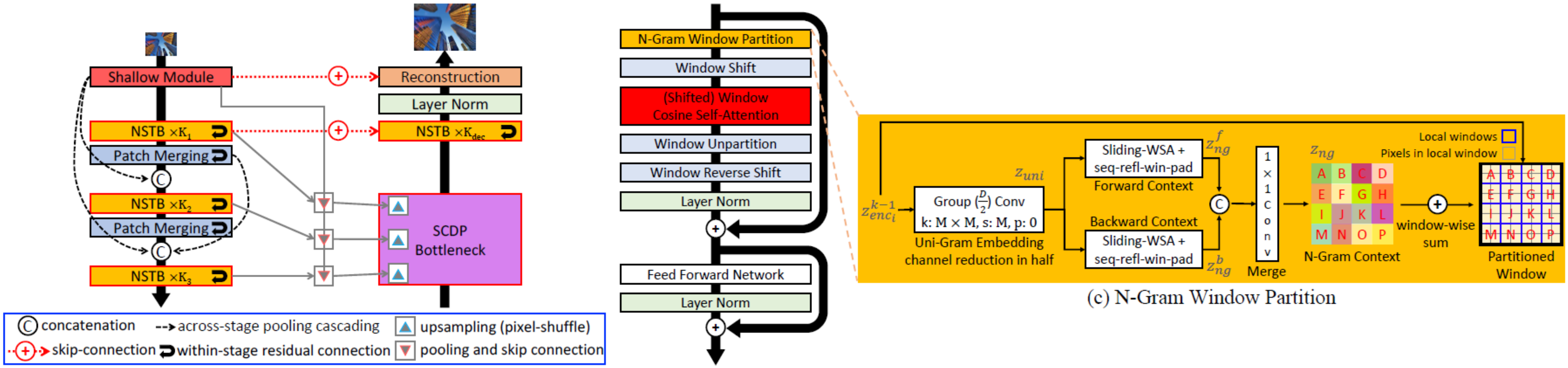
❖ Lightweight SR Results (SwinIR-NG = SwinIR-light + N-Gram)

Method	Year	Scale	Mult-Adds	#Params	Set5		Set14		BSD100		Urban100		Manga109	
					PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SwinIR-light [38]	2021	×2	243.7G	910K	38.14	0.9611	33.86	0.9206	32.31	0.9012	32.76	0.9340	39.12	0.9783
ESRT [48]	2022	×2	191.4G	677K	38.03	0.9600	33.75	0.9184	32.25	0.9001	32.58	0.9318	39.12	0.9774
ELAN-light [79]	2022	×2	168.4G	582K	38.17	0.9611	33.94	0.9207	32.30	0.9012	32.76	0.9340	39.12	0.9783
DiVANet [7]	2023	×2	189.0G	902K	38.16	0.9612	33.80	0.9195	32.29	0.9012	32.60	0.9325	39.08	0.9775
SwinIR-NG (ours)	**2023**	×2	**274.1G**	**1,181K**	**38.17**	**0.9612**	**33.94**	**0.9205**	**32.31**	**0.9013**	**32.78**	**0.9340**	**39.20**	**0.9781**
SwinIR-light [38]	2021	×3	109.5G	918K	34.62	0.9289	30.54	0.8463	29.20	0.8082	28.66	0.8624	33.98	0.9478
ESRT [48]	2022	×3	96.4G	770K	34.42	0.9268	30.43	0.8433	29.15	0.8063	28.46	0.8574	33.95	0.9455
ELAN-light [79]	2022	×3	75.7G	590K	34.61	0.9288	30.55	0.8463	29.21	0.8081	28.69	0.8624	34.00	0.9478
DiVANet [7]	2023	×3	89.0G	949K	34.60	0.9285	30.47	0.8447	29.19	0.8073	28.58	0.8603	33.94	0.9468
SwinIR-NG (ours)	**2023**	×3	**114.1G**	**1,190K**	**34.64**	**0.9293**	**30.58**	**0.8471**	**29.24**	**0.8090**	**28.75**	**0.8639**	**34.22**	**0.9488**
SwinIR-light [38]	2021	×4	61.7G	930K	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
ESRT [48]	2022	×4	67.7G	751K	32.19	0.8947	28.69	0.7833	27.69	0.7379	26.39	0.7962	30.75	0.9100
ELAN-light [79]	2022	×4	43.2G	601K	32.43	0.8975	28.78	0.7858	27.69	0.7406	26.54	0.7982	30.92	0.9150
DiVANet [7]	2023	×4	57.0G	939K	32.41	0.8973	28.70	0.7844	27.65	0.7391	26.42	0.7958	30.73	0.9119
SwinIR-NG↓ (ours)	**2023**	×4	**42.5G**	**770K**	**32.44**	**0.8978**	**28.80**	**0.7863**	**27.70**	**0.7407**	**26.47**	**0.7977**	**30.97**	**0.9147**
SwinIR-NG↓§ (ours)			**42.5G**	**770K**	**32.48**	**0.8979**	**28.83**	**0.7868**	**27.71**	**0.7411**	**26.54**	**0.7998**	**31.12**	**0.9158**
SwinIR-NG (ours)			**63.0G**	**1,201K**	**32.44**	**0.8980**	**28.83**	**0.7870**	**27.73**	**0.7418**	**26.61**	**0.8010**	**31.09**	**0.9161**

❖ Visual Comparisons



1 Methodology – Overall Architecture



(a) Overall architecture of NGswin

(b) NSTB (proposed)

(c) N-Gram Window Partition

Core Design

- (1.2) NSTB (N-Gram Swin Transformer Block): N-Gram Context
- (1.3) SCDP Bottleneck

1.1 Methodology – N-Gram Context

N-Gram Context Motivation

NLP, Text analysis
Transformer

→

Vision, Image Processing
Vision Transformer

NLP, Text analysis
??????

⇢

Vision, Image Processing
Swin Transformer

☞ ☞ ☞ N-Gram Context

1.1 Methodology – N-Gram Context

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

SOGANG UNIVERSITY Machine Learning Laboratory

N-Gram Context in Text Analysis

Sentence1: We use laptop pc for office work.

→N-Gram (forward): (SOS we), (we use), (use laptop), (laptop pc), (pc for), (for office), (office work), (work EOS)

→N-Gram (backward): (SOS we), (we use), (use laptop), (laptop pc), (pc for), (for office), (office work), (work EOS)

→N-Gram (bi-directional): (SOS we use), (we use laptop), (use laptop pc), (laptop pc for), (pc for office),
(for office work), (office, work, EOS)

*SOS: Start Of Sentence
*EOS: End Of Sentence

Sentence2: Office work makes me tired.

→N-Gram (forward): (SOS office), (office work), (work makes), (makes me), (me tired), (tired EOS)

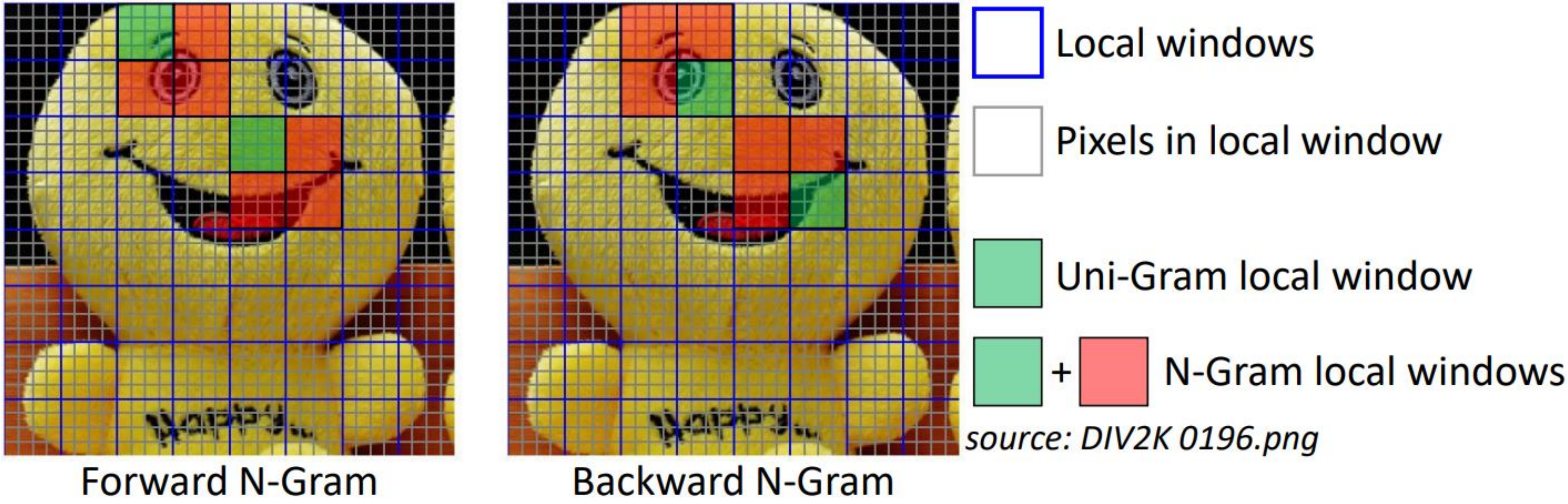
→N-Gram (backward): (SOS office), (office work), (work makes), (makes me), (me tired), (tired EOS)

→N-Gram (bi-directional): (SOS office work), (office work makes), (work makes me), (makes me tired),
(me tired EOS)

- Character: an alphabet in each word
- Uni-Gram: a word in each sentence
- N-Gram: a word pair neighboring Uni-Gram

1.1 Methodology – N-Gram Context

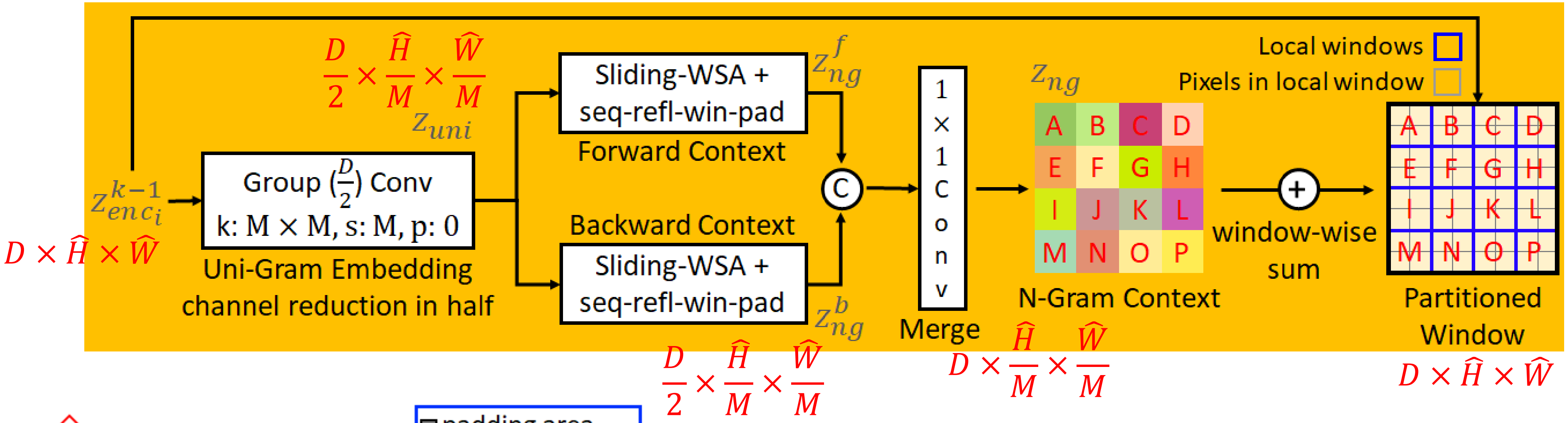
Definition of N-Gram in an Image for Swin Transformer

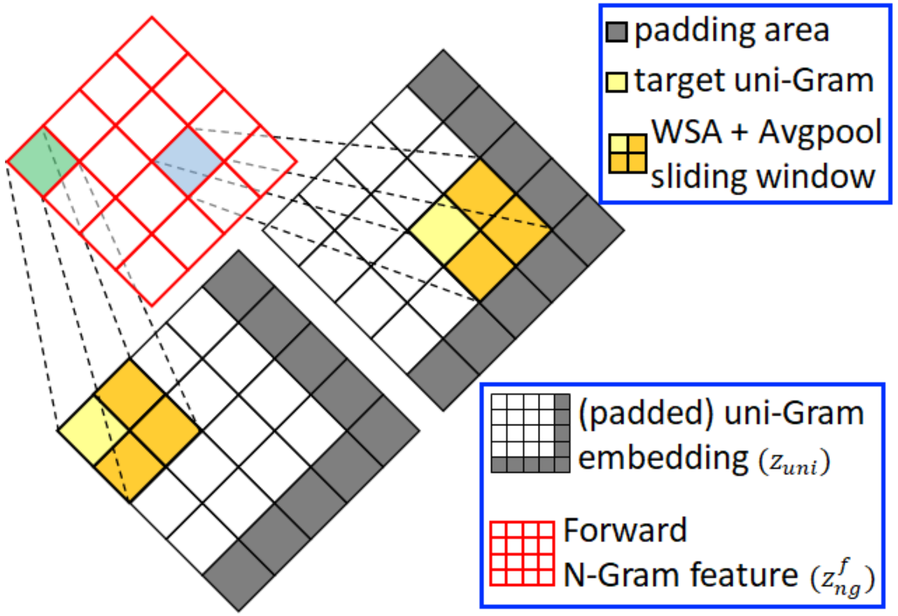


- Character: Alphabet – Pixel
- Uni-Gram: Word – Window
- N-Gram: Neighbor Words Pair – Neighbor Windows Set

1.1 Methodology – N-Gram Context

Architecture of N-Gram Window Partitioning





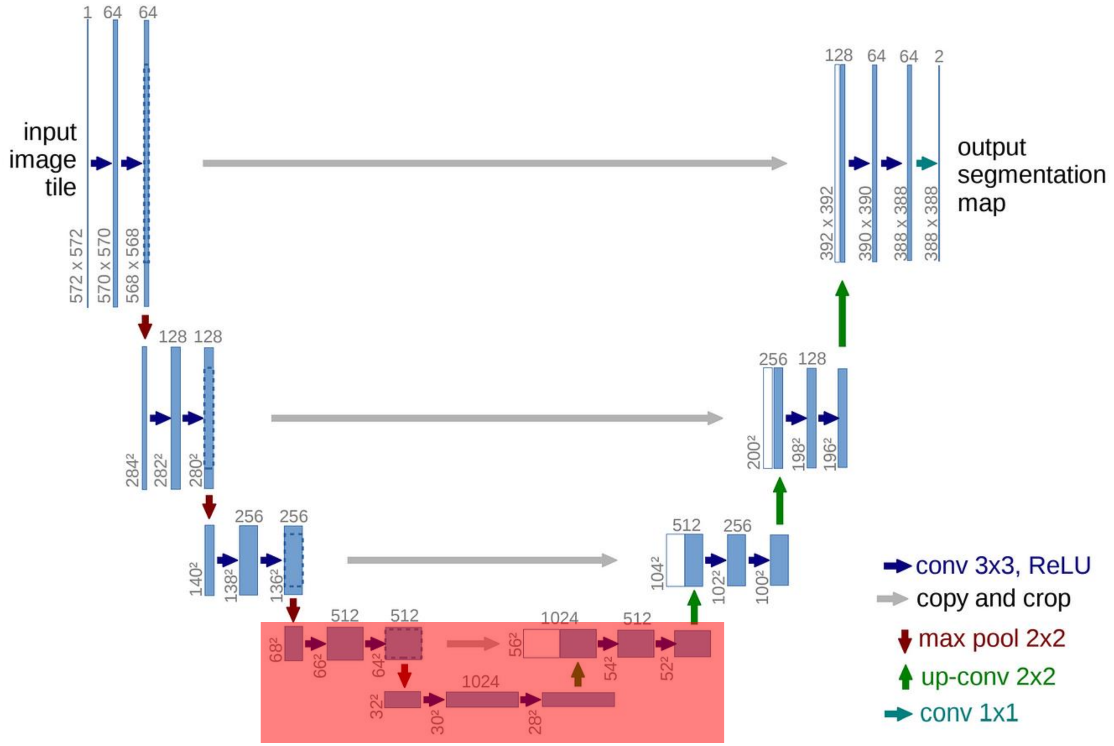
Sliding-WSA
(Window Self-Attention)

1.2 Methodology – SCDP Bottleneck

SCDP (pixel Shuffle – Concatenation – Depth-wise conv – Point-wise conv

Problem of Standard U-Net Based Bottleneck

- The resolution of bottleneck input is too low.
- Correspondingly, the next layer also takes low-resolution features.
- However, successful super-resolution tasks depend on how many the network handles high-resolution feature maps.
- Nevertheless, a hierarchical U-Net structure is more efficient (Table below).



Standard U-Net [1] Architecture

Table 1. Comparison of computational complexity with state-of-the-art networks. Our NGswin is much more efficient. Mult-Adds is evaluated on a 1280 × 720 HR image.

Scale	NGswin	SwinIR-light [38][2]	ESRT [48]	DiVANet [7]	ELAN-light [79]
x2	**140.4G**	243.7G	191.4G	189.0G	168.4G
x3	**66.6G**	109.5G	96.4G	89.0G	75.7G
x4	**36.4G**	61.7G	67.7G	57.0G	43.2G

[1] U-Net: Convolutional Networks for Biomedical Image Segmentation – Olaf Ronneberger et al. (MICCAI15)

1.2 Methodology – SCDP Bottleneck

SCDP (pixel Shuffle – Concatenation – Depth-wise conv – Point-wise conv

Algorithm 1 SCDP Bottleneck Pseudo-code, PyTorch-like

```
# zi: output list of last NSTBs in three encoder stages
# zs: output of shallow module

x = list()
for i in range(3): # pixel-"S"huffle
    x_ = zi[i] + down(zs, i) # before shuffling
    x.append(PixelShuffle(x_, 2**i))
x = torch.cat(x, dim=-1) # "C"oncatenation
x = Rearrange(x, '(h w) d -> d h w') # ignores batch
x = GELU(depth_wise(x)) # "D"epth-wise convolution
x = Rearrange(x, 'd h w -> (h w) d')
x = LayerNorm(point_wise(x)) # "P"oint-wise projection

def down(z, exp): # downsizing zs
    z = Rearrange(z, '(h w) d -> d h w')
    for e in range(exp): # iterative max-poolings
        z = MaxPool2D(z) # 2x2 pool
    z = LeakyReLU(z)
    return Rearrange(z, 'd h w -> (h w) d')
```

Table 6. Ablation study on extra stages and SCDP bottleneck.

(a) The specifications of models with different stages. dep.: # of NSTBs / res.: training input resolution. The total number of NSTBs is kept as 20.

Stages	encoder1	encoder2	encoder3	encoder4	decoder1	decoder2
	dep. / res.	dep. / res.	dep. / res.	dep. / res.	dep. / res.	dep. / res.
extra	4 / 64×64	4 / 32×32	4 / 16×16	4 / 8×8	2 / 32×32	2 / 64×64
default	**6 / 64×64**	**4 / 32×32**	**4 / 16×16**	**- / -**	**6 / 64×64**	**- / -**

(b) Impacts of extra stages and SCDP bottleneck. PSNR / SSIM.

Stages	SCDP	Scale	Mult-Adds	#Params	Urban100	Manga109
extra	*w/o*		87.98G	997K	32.28 / 0.9298	38.72 / 0.9773
default	*w/o*	×2	138.88G	992K	32.48 / 0.9321	38.92 / 0.9776
default	***w/***		**140.41G**	**998K**	**32.53 / 0.9324**	**38.97 / 0.9777**
extra	*w/o*		42.10G	1,006K	28.33 / 0.8562	33.67 / 0.9453
default	*w/o*	×3	65.85G	1,001K	28.47 / 0.8596	33.81 / 0.9464
default	***w/***		**66.56G**	**1,007K**	**28.52 / 0.8603**	**33.89 / 0.9470**
extra	*w/o*		23.33G	1,018K	26.22 / 0.7900	30.46 / 0.9090
default	*w/o*	×4	36.06G	1,013K	26.38 / 0.7954	30.71 / 0.9121
default	***w/***		**36.44G**	**1,019K**	**26.45 / 0.7963**	**30.80 / 0.9128**

3 Results

Main Results

Method	Training Dataset	Scale	Mult-Adds	#Params	Set5 PSNR	Set5 SSIM	Set14 PSNR	Set14 SSIM	BSD100 PSNR	BSD100 SSIM	Urban100 PSNR	Urban100 SSIM	Manga109 PSNR	Manga109 SSIM
EDSR-baseline [28]	D2K	×2	316.3G	1,370K	37.99	0.9604	33.57	0.9175	32.16	0.8994	31.98	0.9272	38.54	0.9769
MemNet [50]	291	×2	2,662.4G	677K	37.78	0.9597	33.28	0.9142	32.08	0.8978	31.31	0.9195	-	-
CARN [2]	D2K+291	×2	222.8G	1,592K	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	38.36	0.9765
IMDN [23]	D2K	×2	158.8G	694K	38.00	0.9605	33.63	0.9177	32.19	0.8996	32.17	0.9283	38.88	0.9774
LatticeNet [37]	D2K	×2	169.5G	756K	38.06	0.9607	33.70	0.9187	32.20	0.8999	32.25	0.9288	38.94	0.9774
RFDN-L [30]	D2K	×2	145.8G	626K	38.08	0.9606	33.67	0.9190	32.18	0.8996	32.24	0.9290	38.95	0.9773
SRPN-Lite [65]	DF2K	×2	139.9G	609K	38.10	0.9608	33.70	0.9189	32.25	0.9005	32.26	0.9294	-	-
HNCT [16]	D2K	×2	82.4G	357K	38.08	0.9608	33.65	0.9182	32.22	0.9001	32.22	0.9294	38.87	0.9774
FMEN [15]	DF2K	×2	172.0G	748K	38.10	0.9609	33.75	0.9192	32.26	0.9007	32.41	0.9311	38.95	0.9778
NGswin (ours)	**D2K**	×2	**140.4G**	**998K**	**38.05**	**0.9610**	**33.79**	**0.9199**	**32.27**	**0.9008**	**32.53**	**0.9324**	**38.97**	**0.9777**
EDSR-baseline [28]	D2K	×3	160.2G	1,555K	34.37	0.9270	30.28	0.8417	29.09	0.8052	28.15	0.8527	33.45	0.9439
MemNet [50]	219	×3	2,662.4G	677K	34.09	0.9248	30.00	0.8350	28.96	0.8001	27.56	0.8376	-	-
CARN [2]	D2K+291	×3	118.8G	1,592K	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	33.50	0.9440
IMDN [23]	D2K	×3	71.5G	703K	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519	33.61	0.9445
LatticeNet [37]	D2K	×3	76.3G	765K	34.40	0.9272	30.32	0.8416	29.10	0.8049	28.19	0.8513	33.63	0.9442
RFDN-L [30]	D2K	×3	65.6G	633K	34.47	0.9280	30.35	0.8421	29.11	0.8053	28.32	0.8547	33.78	0.9458
SRPN-Lite [65]	DF2K	×3	62.7G	615K	34.47	0.9276	30.38	0.8425	29.16	0.8061	28.22	0.8534	-	-
HNCT [16]	D2K	×3	37.8G	363K	34.47	0.9275	30.44	0.8439	29.15	0.8067	28.28	0.8557	33.81	0.9459
FMEN [15]	DF2K	×3	77.2G	757K	34.45	0.9275	30.40	0.8435	29.17	0.8063	28.33	0.8562	33.86	0.9462
NGswin (ours)	**D2K**	×3	**66.6G**	**1,007K**	**34.52**	**0.9282**	**30.53**	**0.8456**	**29.19**	**0.8078**	**28.52**	**0.8603**	**33.89**	**0.9470**
EDSR-baseline [28]	D2K	×4	114.0G	1,518K	32.09	0.8938	28.58	0.7813	27.57	0.7357	26.04	0.7849	30.35	0.9067
MemNet [50]	291	×4	2,662.4G	677K	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630	-	-
CARN [2]	D2K+291	×4	90.9G	1,592K	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.47	0.9084
IMDN [23]	D2K	×4	40.9G	715K	32.21	0.8948	28.58	0.7811	27.56	0.7353	26.04	0.7838	30.45	0.9075
LatticeNet [37]	D2K	×4	43.6G	777K	32.18	0.8943	28.61	0.7812	27.57	0.7355	26.14	0.7844	30.54	0.9075
RFDN-L [30]	D2K	×4	37.4G	643K	32.28	0.8957	28.61	0.7818	27.58	0.7363	26.20	0.7883	30.61	0.9096
SRPN-Lite [65]	DF2K	×4	35.8G	623K	32.24	0.8958	28.69	0.7836	27.63	0.7373	26.16	0.7875	-	-
HNCT [16]	D2K	×4	22.0G	373K	32.31	0.8957	28.71	0.7834	27.63	0.7381	26.20	0.7896	30.70	0.9112
FMEN [15]	DF2K	×4	44.2G	769K	32.24	0.8955	28.70	0.7839	27.63	0.7379	26.28	0.7908	30.70	0.9107
NGswin (ours)	**D2K**	×4	**36.4G**	**1,019K**	**32.33**	**0.8963**	**28.78**	**0.7859**	**27.66**	**0.7396**	**26.45**	**0.7963**	**30.80**	**0.9128**

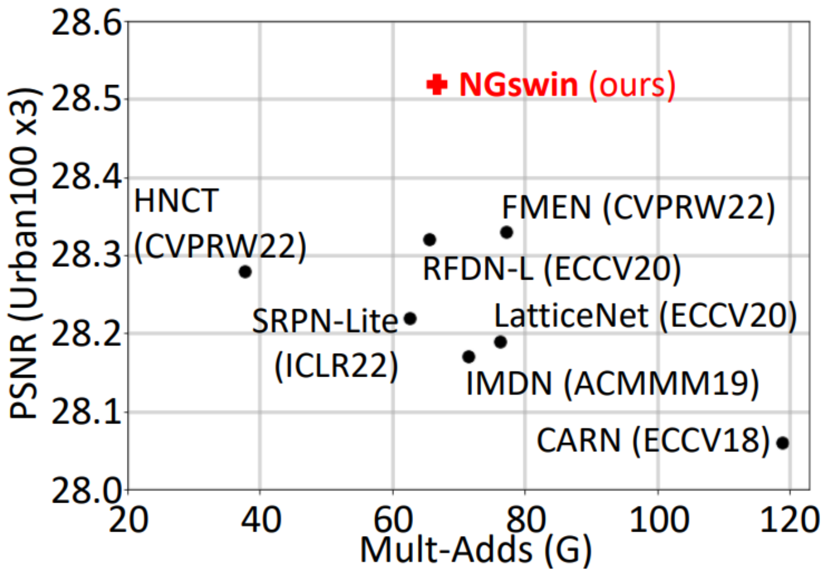
Method	Year	Scale	Mult-Adds	#Params	Set5 PSNR	Set5 SSIM	Set14 PSNR	Set14 SSIM	BSD100 PSNR	BSD100 SSIM	Urban100 PSNR	Urban100 SSIM	Manga109 PSNR	Manga109 SSIM
SwinIR-light [27]	2021	×2	243.7G	910K	38.14	0.9611	33.86	0.9206	32.31	0.9012	32.76	0.9340	39.12	0.9783
ESRT [35]	2022	×2	191.4G	677K	38.03	0.9600	33.75	0.9184	32.25	0.9001	32.58	0.9318	39.12	0.9774
ELAN-light [63]	2022	×2	168.4G	582K	38.17	0.9611	33.94	0.9207	32.30	0.9012	32.76	0.9340	39.12	0.9783
DiVANet [6]	2023	×2	189.0G	902K	38.16	0.9612	33.80	0.9195	32.29	0.9012	32.60	0.9325	39.08	0.9775
SwinIR-NG (ours)	**2023**	×2	**274.1G**	**1,181K**	**38.17**	**0.9612**	**33.94**	**0.9205**	**32.31**	**0.9013**	**32.78**	**0.9340**	**39.20**	**0.9781**
SwinIR-light [27]	2021	×3	109.5G	918K	34.62	0.9289	30.54	0.8463	29.20	0.8082	28.66	0.8624	33.98	0.9478
ESRT [35]	2022	×3	96.4G	770K	34.42	0.9268	30.43	0.8433	29.15	0.8063	28.46	0.8574	33.95	0.9455
ELAN-light [63]	2022	×3	75.7G	590K	34.61	0.9288	30.55	0.8463	29.21	0.8081	28.69	0.8624	34.00	0.9478
DiVANet [6]	2023	×3	89.0G	949K	34.60	0.9285	30.47	0.8447	29.19	0.8073	28.58	0.8603	33.94	0.9468
SwinIR-NG (ours)	**2023**	×3	**114.1G**	**1,190K**	**34.64**	**0.9293**	**30.58**	**0.8471**	**29.24**	**0.8090**	**28.75**	**0.8639**	**34.22**	**0.9488**
SwinIR-light [27]	2021	×4	61.7G	930K	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
ESRT [35]	2022	×4	67.7G	751K	32.19	0.8947	28.69	0.7833	27.69	0.7379	26.39	0.7962	30.75	0.9100
ELAN-light [63]	2022	×4	43.2G	601K	32.43	0.8975	28.78	0.7858	27.69	0.7406	26.54	0.7982	30.92	0.9150
DiVANet [6]	2023	×4	57.0G	939K	32.41	0.8973	28.70	0.7844	27.65	0.7391	26.42	0.7958	30.73	0.9119
SwinIR-NG↓ (ours)			**42.5G**	**770K**	**32.44**	**0.8978**	**28.80**	**0.7863**	**27.70**	**0.7407**	**26.47**	**0.7977**	**30.97**	**0.9147**
SwinIR-NG↓§ (ours)	**2023**	×4	**42.5G**	**770K**	**32.48**	**0.8979**	**28.83**	**0.7868**	**27.71**	**0.7411**	**26.54**	**0.7998**	**31.12**	**0.9158**
SwinIR-NG (ours)			**63.0G**	**1,201K**	**32.44**	**0.8980**	**28.83**	**0.7870**	**27.73**	**0.7418**	**26.61**	**0.8010**	**31.09**	**0.9161**

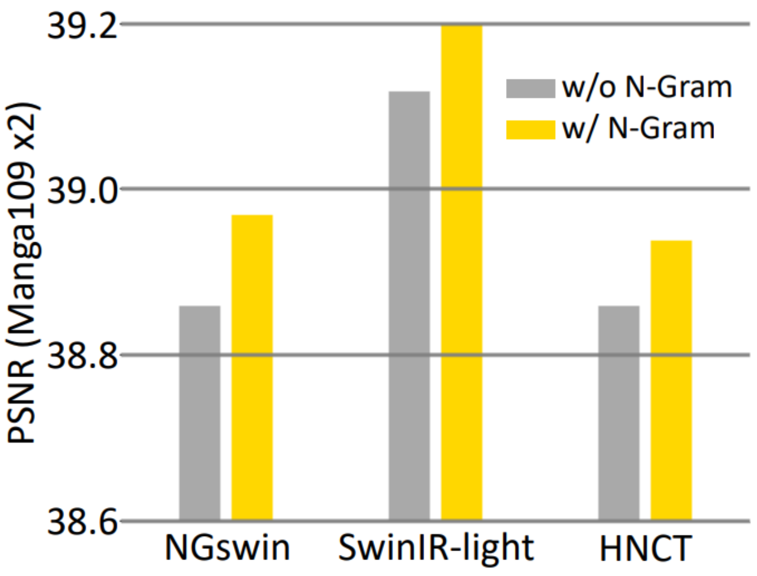
1st Track: Efficient Super-Resolution (NGswin)

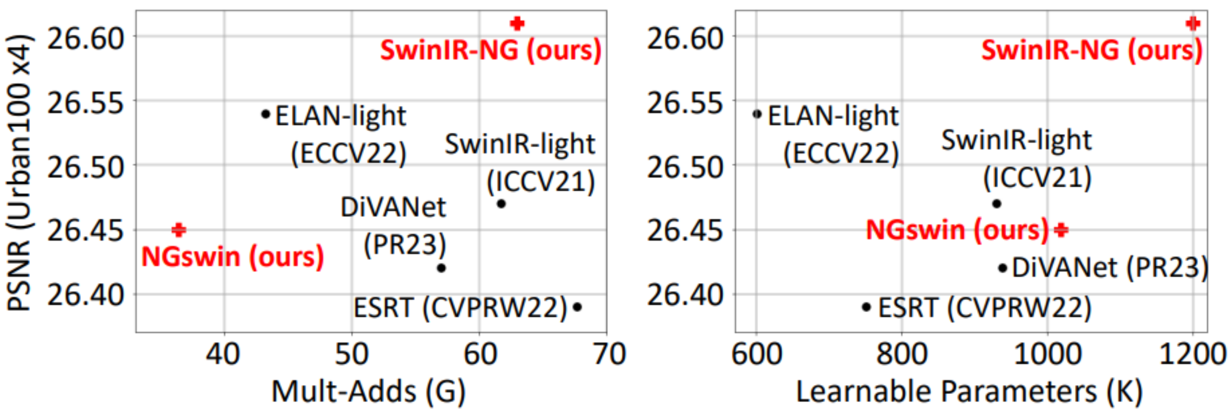
2nd Track: Lightweight Super-Resolution (SwinIR-NG = SwinIR-light + N-Gram)

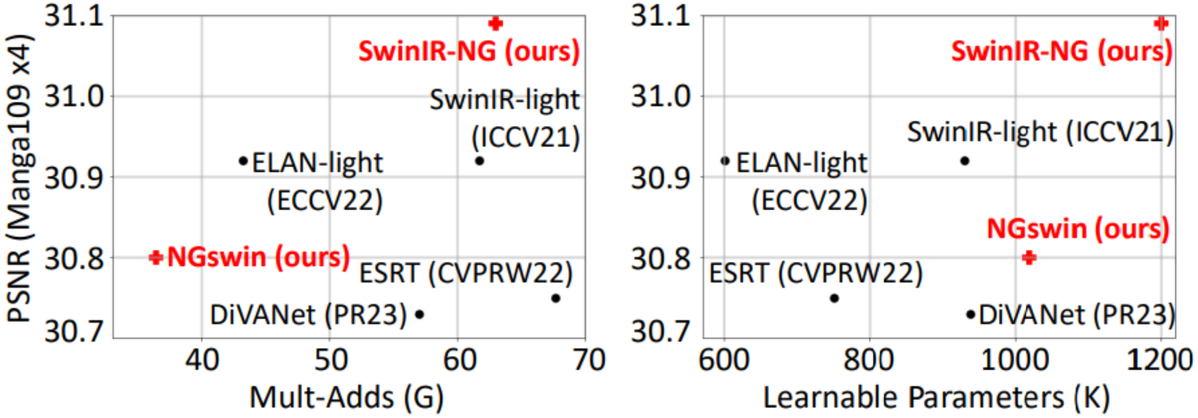
3 Results

Main Results









3 Results

Ablations

(a) N-Gram context (Tab. 4).

NGswin without *vs.* with N-Gram

N-Gram	Scale	Mult-Adds	#Params	Set5	Set14	BSD100	Urban100	Manga109
w/o	×2	138.20G	750K	**38.05** / 0.9609	33.70 / 0.9194	32.25 / 0.9006	32.39 / 0.9304	38.86 / 0.9775
w/	×2	**140.41G**	**998K**	**38.05 / 0.9610**	**33.79 / 0.9199**	**32.27 / 0.9008**	**32.53 / 0.9324**	**38.97 / 0.9777**
w/o	×3	65.53G	759K	**34.53** / 0.9281	30.48 / 0.8451	29.15 / 0.8073	28.37 / 0.8573	33.81 / 0.9464
w/	×3	**66.56G**	**1,007K**	34.52 / **0.9282**	**30.53 / 0.8456**	**29.19 / 0.8078**	**28.52 / 0.8603**	**33.89 / 0.9470**
w/o	×4	35.89G	771K	32.34 / 0.8963	28.70 / 0.7844	27.63 / 0.7390	26.25 / 0.7918	30.70 / 0.9123
w/o (channel up)	×4	53.71G	1,189K	32.37 / **0.8973**	28.75 / 0.7854	27.65 / 0.7396	26.28 / 0.7927	30.73 / 0.9129
w/o (depth up)	×4	47.88G	1,061K	**32.40** / 0.8967	28.75 / 0.7853	**27.66 / 0.7398**	26.37 / 0.7946	30.78 / **0.9133**
w/	×4	**36.44G**	**1,019K**	32.33 / 0.8963	**28.78 / 0.7859**	**27.66** / 0.7396	**26.45 / 0.7963**	**30.80** / 0.9128

HNCT *vs.* HNCT-NG

N-Gram	Scale	Mult-Adds	#Params	Set5	Set14	BSD100	Urban100	Manga109
w/o	×2	82.39G	357K	38.08 / 0.9608	**33.65** / 0.9182	32.22 / 0.9001	32.22 / 0.9294	38.87 / **0.9774**
w/	×2	**83.19G**	**424K**	**38.10 / 0.9610**	33.64 / **0.9195**	**32.25 / 0.9006**	**32.35 / 0.9306**	**38.94 / 0.9774**
w/o	×3	37.78G	363K	34.47 / 0.9275	30.44 / 0.8439	29.15 / 0.8067	28.28 / 0.8557	**33.81** / 0.9459
w/	×3	**38.14G**	**431K**	**34.48 / 0.9280**	**30.48 / 0.8450**	**29.16 / 0.8074**	**28.38 / 0.8573**	**33.81 / 0.9464**
w/o	×4	22.01G	373K	32.31 / 0.8957	28.71 / 0.7834	27.63 / 0.7381	26.20 / 0.7896	30.70 / 0.9112
w/	×4	**22.21G**	**440K**	**32.32 / 0.8960**	**28.72 / 0.7846**	**27.65 / 0.7391**	**26.23 / 0.7912**	**30.71 / 0.9114**

(b) N-Gram directions and interaction (Tab. 5). The second best results are in underline.

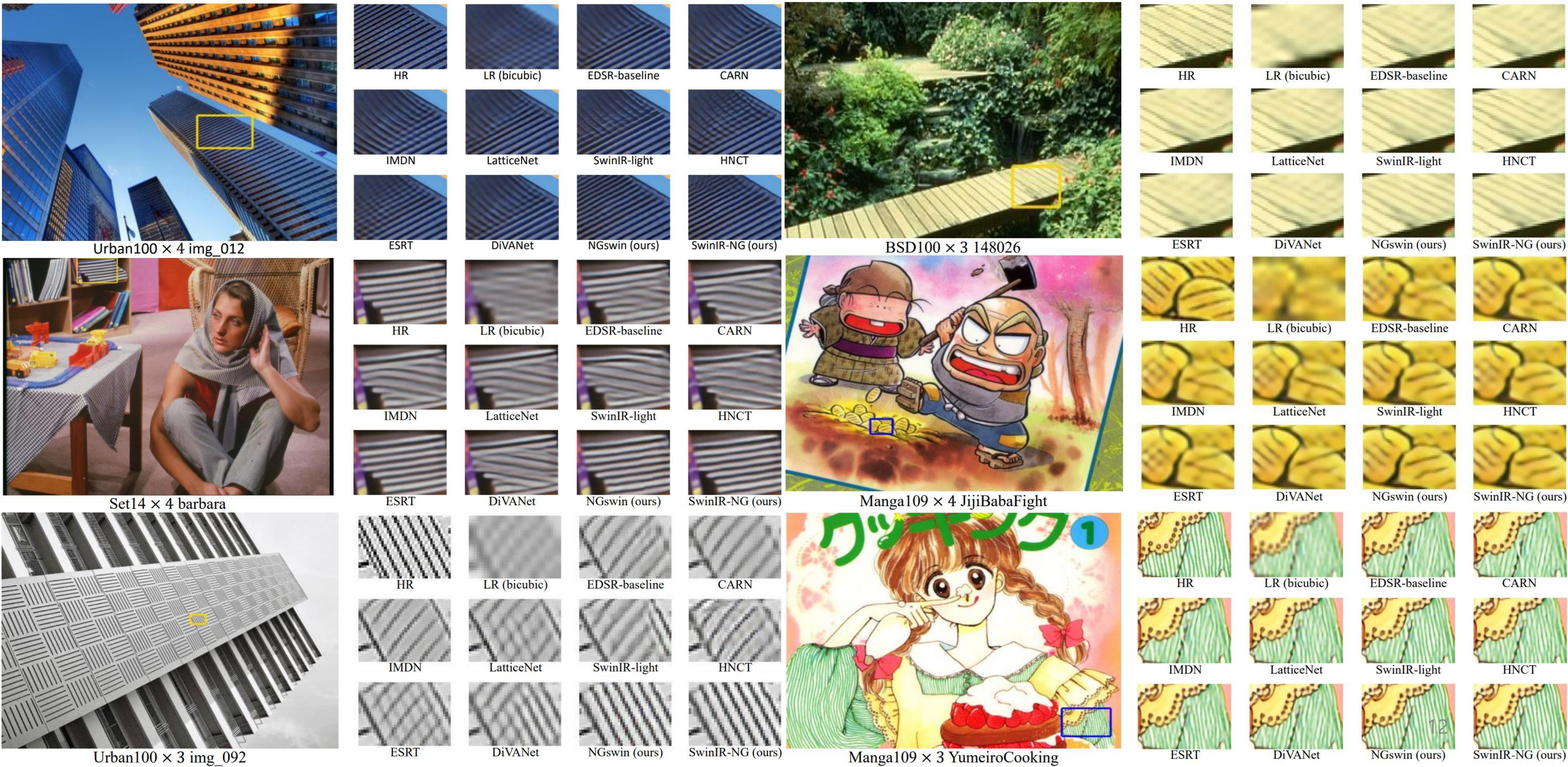
Direction	Type	Mult-Adds	#Params	Set5	Set14	BSD100	Urban100	Manga109
1	WSA	152.41G	1,238,056	38.05 / **0.9610**	33.78 / 0.9198	32.26 / 0.9006	**32.54** / 0.9322	38.90 / **0.9777**
4	WSA	139.56G	935,272	**38.07** / 0.9609	33.76 / 0.9197	32.25 / 0.9007	32.52 / 0.9317	38.92 / 0.9776
1	CNN	139.80G	1,327,528	38.04 / **0.9610**	33.77 / 0.9197	32.25 / 0.9005	32.45 / 0.9316	38.86 / 0.9775
2	CNN	139.38G	998,568	38.04 / **0.9610**	**33.83 / 0.9203**	32.26 / 0.9007	**32.54** / 0.9321	38.90 / 0.9776
4	CNN	139.17G	936,488	38.02 / 0.9609	33.77 / 0.9178	32.26 / 0.9006	32.52 / 0.9320	38.93 / **0.9777**
2	**WSA**	**140.41G**	**998,384**	38.05 / **0.9610**	33.79 / 0.9199	**32.27 / 0.9008**	32.53 / **0.9324**	**38.97 / 0.9777**

3 Results

Visual Results (*vs.* other networks)

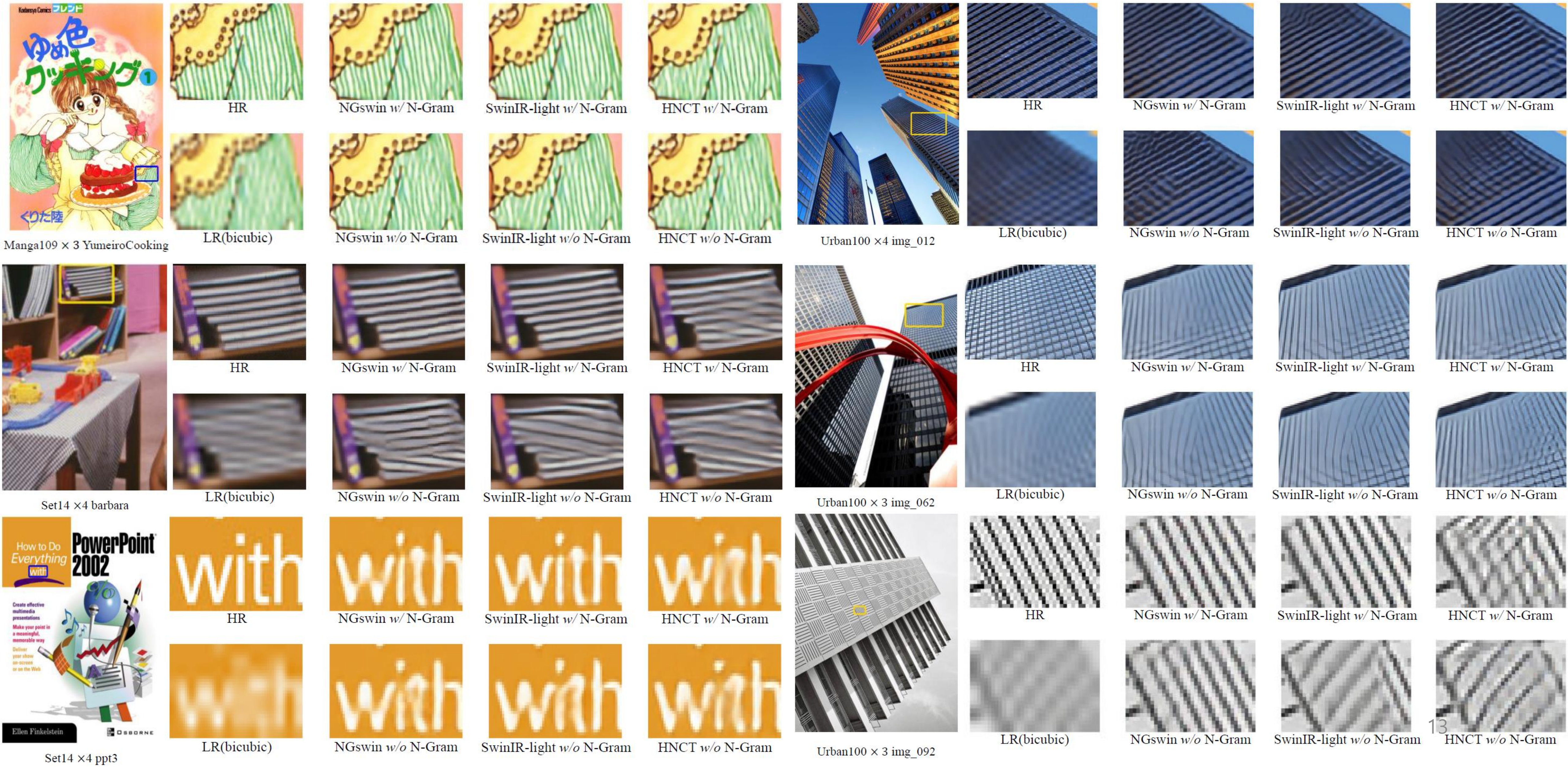
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

SOGANG UNIVERSITY Machine Learning Laboratory



3 Results

Visual Results (*w/o* N-Gram *vs.* *w/* N-Gram)



Thank You