# NTIRE 2023 ISR

Test case defence

# Main goals

**I. Overview of NTIRE 2023 Image Super Resolution Competition:**

- Used Train and Test data overview
- An overview of the most interesting models from the competition
  - Architectures | Objectives | Train strategies

**II. Chosen models inference:**

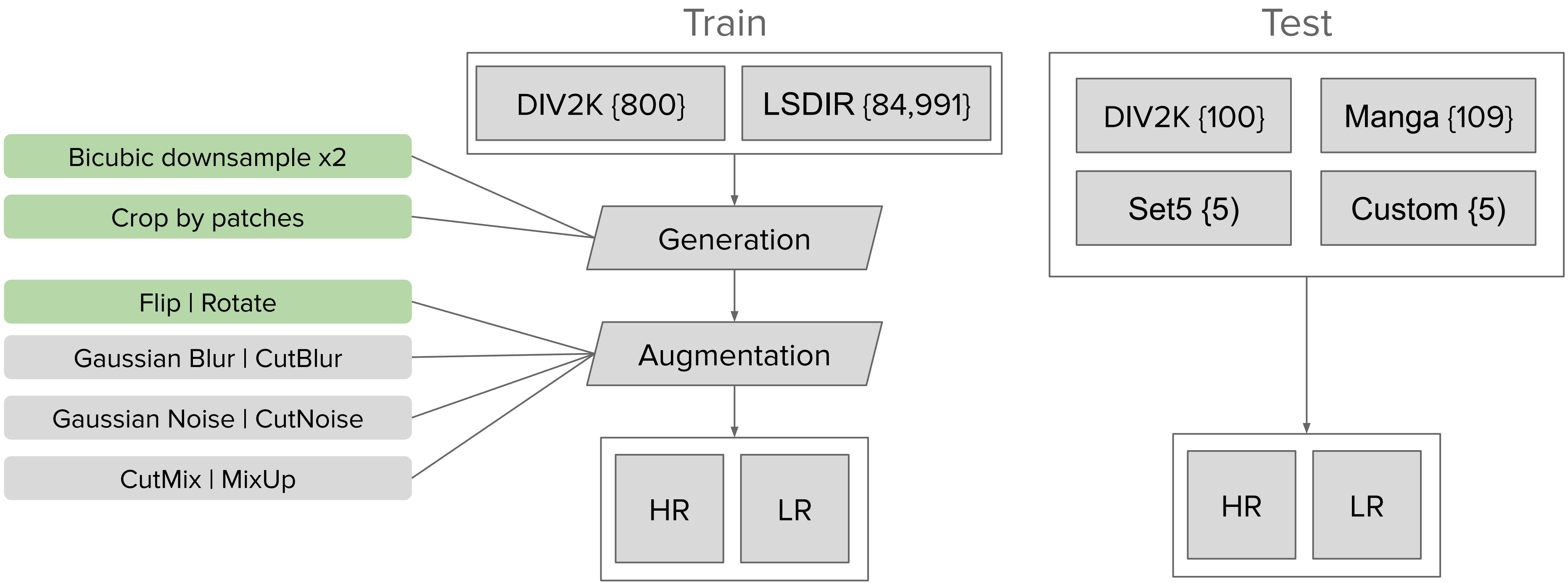
- Check quality of chosen models inference on custom dataset or datasets

**III. Models training for additional tasks:**

- Denoising
- Deblurring

**IV. How to improve the perceptual quality of generated image by loss modifications**

# Dataset overview



# Models architectures overview. MDRN

## I. Multi-level Dispersion Residual Network (MDRN)

NTIRE 2023 winner under:

- Number of parameters
- FLOPS

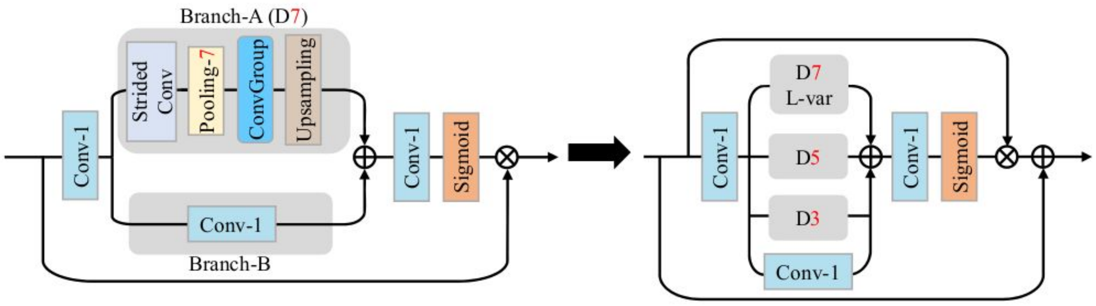
Architecture key aspects:

Enhanced spatial attention (ESA)

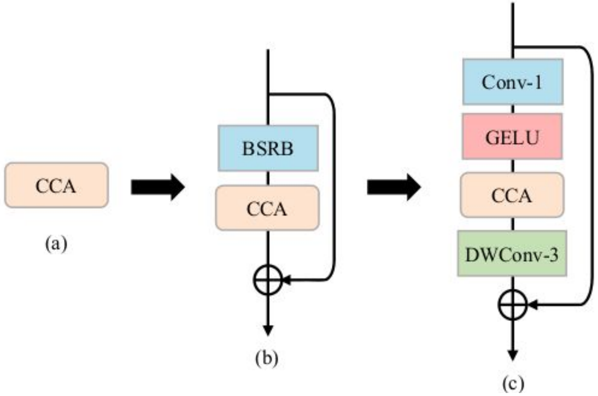
Multi-level Dispersion Spatial Attention (MDSA)

Contrast-aware channel attention (CCA)

Enhanced Contrast-aware Channel Attention (ECCA)




ESA -> MDSA




CCA -> ECCA

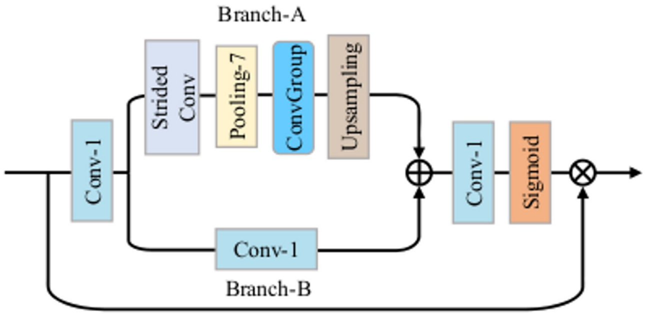
# MDRN. ESA block

**ESA** consists of **two parts**: **Branch-A** and **Branch-B**.

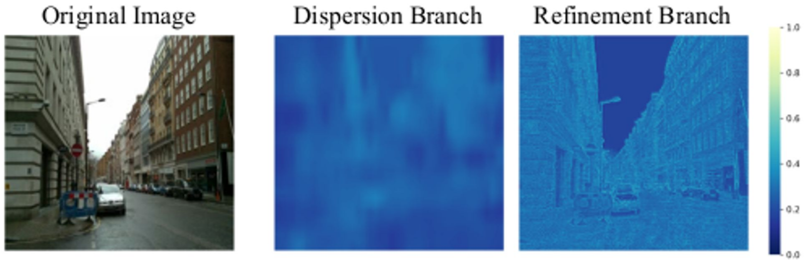
**Activate both** branches separately by the **sigmoid function** and visualize the results as shown in Fig.

**Branch-A** was mainly used to **generate fine attention map in original resolution space**.

**Branch-B** was used to **generate dispersion attention map in lower resolution space** because of feature spatial compression and attention weight dispersion

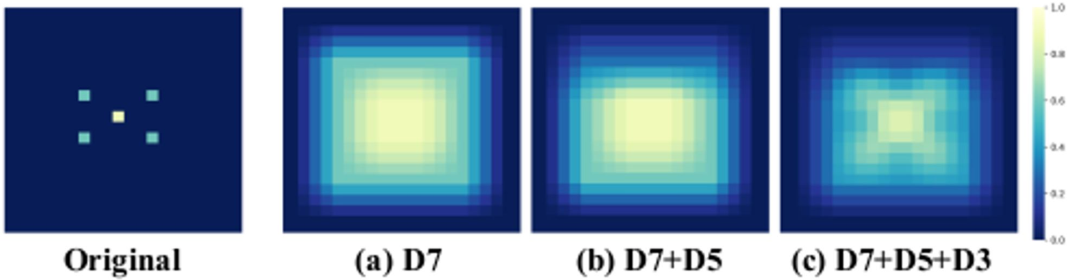


(a) Architecture of ehanced spatial attention (ESA).



(b) Visualization results of different branches in ESA.

# MDRN. MDSA vs ESA difference




**ESA only uses a single large-size pooling kernel in the dispersion branch**, as shown in (a), surrounding **unimportant information** can easily be given **high weights**, and areas that contain true key information cannot be paid attention to.

Based on the thought of coarse-to-fine, **authors extend the single-level dispersion branch** to **multiple levels** to **focus the dispersion attention scores on key areas**, making such areas gradually more prominent, as shown in (b) and (c).

# MDRN. EADB block

Multi-level dispersion spatial attention (**MDSA**) - introduces **multi-scale** and **variance** information to obtain more accurate **spatial attention distribution.**

Enhanced contrast-aware channel attention (**ECCA**) - combines **lightweight convolution layers** and **residual connections** to improve the **efficiency of channel attention.**

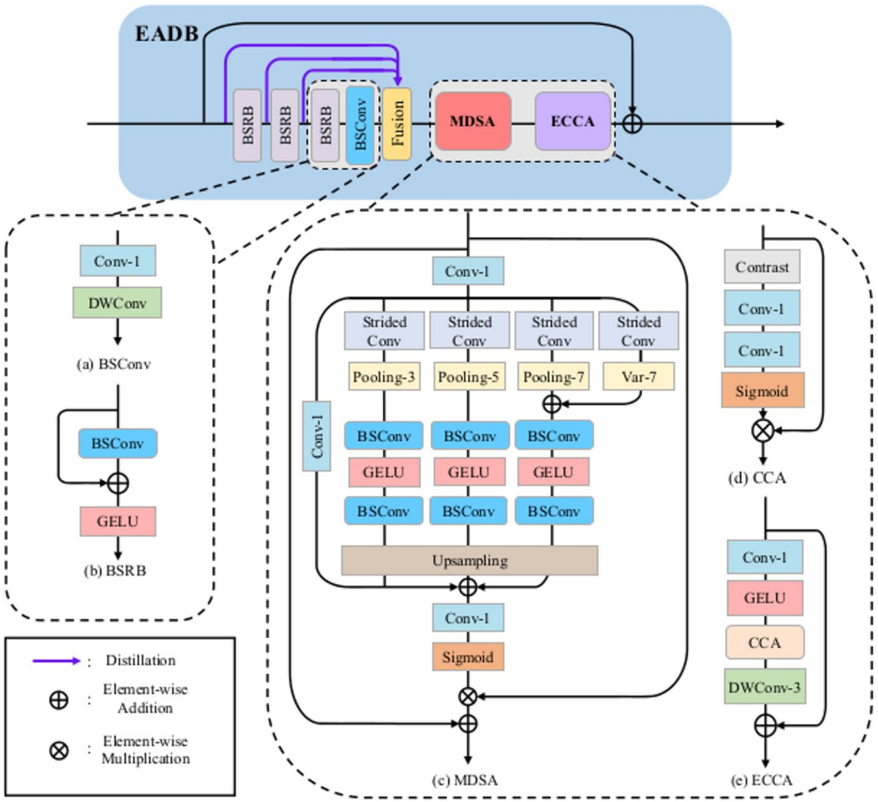



Figure 6. The specific architecture of the proposed enhanced attention distillation block (EADB). (a) BSConv: blueprint separable convolution. 'DWConv' means depth-wise convolution. (b) BSRB: Blueprint Shallow Residual Block. (c) MDSA: multi-level dispersion spatial attention. (d) CCA: contrast-aware channel attention. (e) ECCA: Enhanced contrast-aware channel attention.

# Models architectures overview. GFMN

## II. Gated feature modulation network (GFMN)

NTIRE 2023 3rd place under:

- Number of parameters
- FLOPS

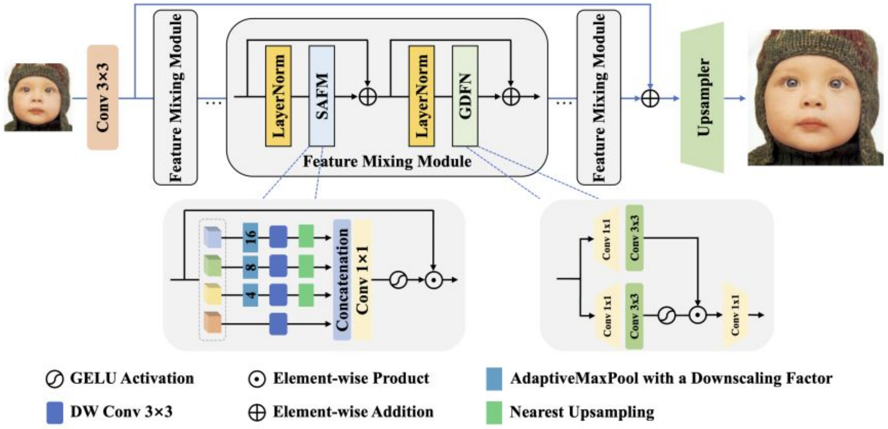
Architecture key aspects:

Convolutional channel mixer (CCM)

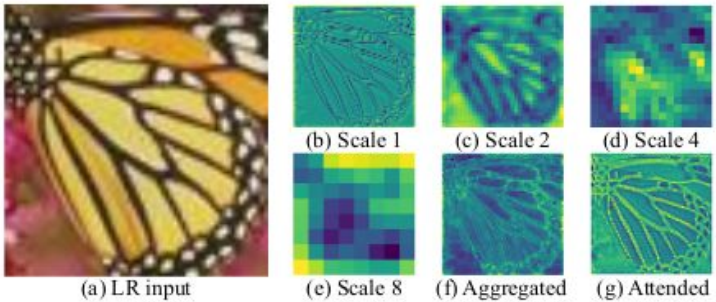
Gated-dconv feed-forward network (GDFN)

Exploiting non-local feature interactions

spatially-adaptive feature modulation (SAFM)




GFMN architecture




Learned deep features from the SAFM module

# Models architectures overview. Comparison

## Multi-level Dispersion Residual Network VS Gated feature modulation network

1. Shallow feature extraction:

Blueprint shallow residual block (BSRB)

3x3 2D Convolution

2. Deep feature extraction:

Enhanced attention distillation block (EADB: BSRB | MDSA | ECCA) **X8**

Feature Mixing Module (FMM: LNorm | SAFM | GDFN) **X8**

3. HR image reconstruction:

3x3 2D Convolution & Pixel Shuffle

3x3 2D Convolution & Pixel Shuffle

| Architecture | Params [M] | FLOPS [G] | Activations [M] |
|---|---|---|---|
| MDRN | 0.095 | 5.58 | 220.88 |
| GFMN | 0.104 | 6.56 | 199.35 |

# Models objectives overview. Train strategies

Multi-level Dispersion Residual Network (MDRN) used train strategy

1. Train **DIV2k 384x384**, batch **64**, **L1** loss minimization, **Adam**, lr **2x1e-3**, 1000k iters.
2. Fine-tune **DIV2K** and 10k **LSDIR**, **384x384**, batch **64**, **Charbonnier** loss minimization, lr **5x1e-4**, 1000k iters.
3. Fine-tune **DIV2k** and 10k **LSDIR**, **480x480**, batch **64**, **L2** loss minimization, lr **2x1e-4**, 650k iters.

Gated feature modulation network (GFMN) used train strategy

1. Train **LSDIR**, **96x96**, batch **64**, **L1 + frequency** losses minimization, **Adam**, lr **1x1e-3** to **1xe-6**, **Cosine Annealing**, 600k iters.

# Models objectives overview. Improvements

How to improve perceptual quality with loss modification?

1. += **L1** loss after **Fourier transform** => to learn **frequency information** (take amplitude and phase).
2. Using **warm strategy** for Fine-tuning.
3. Using **Edge-Aware Loss**: edge detection => L1 between edges.

How to decrease model complexity with architecture modifications?

1. Apply online **convolutional re-parameterization**: complex blocks -> single Convolution layer

# Demo results

## GFMN

| | | |
|---|---|---|
| Set5: | PSNR: 38.0193 | SSIM: 0.9606 |
| Manga109: | PSNR: 30.9854 | SSIM: 0.9386 |
| Christmas: | PSNR: 28.7114 | SSIM: 0.8325 |

2x

2x


**X8 zoom**

## MDRN

| | | |
|---|---|---|
| Set5: | PSNR: 38.1096 | SSIM: 0.9579 |
| Manga109: | PSNR: 30.9723 | SSIM: 0.9390 |
| Christmas: | PSNR: 28.5712 | SSIM: 0.8304 |

2x

2x

# Train results overview. MDRN

# Train results overview. GFMN

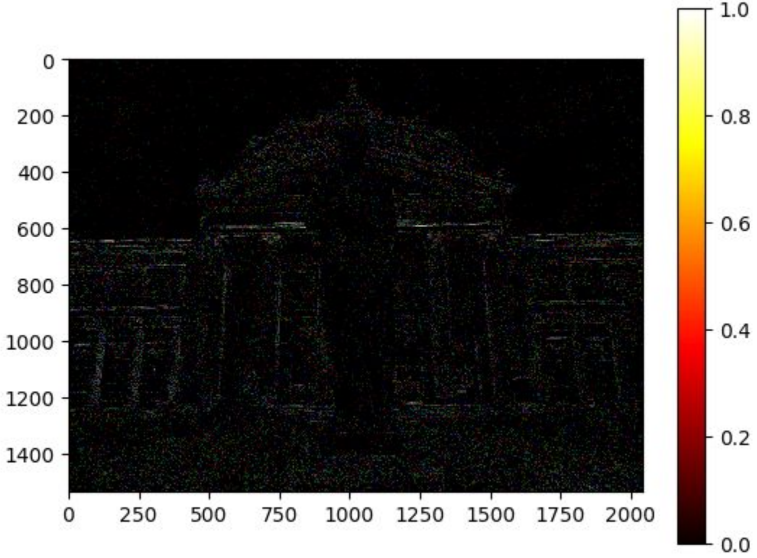
# Train results overview. Metrics

| PSNR metric | DIV2k Blur \| Noise | Manga109 Blur \| Noise | Set5 Blur \| Noise | Christmas Blur \| Noise |
|---|---|---|---|---|
| **MDRN base** | 27.1961 \| 21.6181 | 25.1616 \| 21.5454 | 26.3798 \| 21.7238 | 24.6929 \| 20.9387 |
| **MDRN fine-tuned** | 33.8330 \| 30.8184 | 30.5301 \| 29.3575 | 34.6285 \| 31.0261 | 28.5631 \| 27.4755 |
| **GFMN (SAFM) base** | 27.1902 \| 21.9179 | 25.1580 \| 22.0246 | 26.3741 \| 22.1374 | 24.6895 \| 21.2976 |
| **GFMN (SAFM) fine-tuned** | 32.7189 \| 30.4988 | 29.6930 \| 28.3714 | 33.1904 \| 30.6769 | 28.0334 \| 27.3378 |

Denoise, Deblur

Fine-tune **DIV2K 240x240**, batch **64**, lr **3e-4** to **1e-6**, shed **Cosine Annealing**, gpu **3090**

# Train results overview. Edge-Aware Loss




MDRN fine-tune with L1 loss with Sobel filter. DIV2K PSNR 37.7188 -> 37.790

# Conclusion

**I. Overview of NTIRE 2023 Image Super Resolution Competition:**

- Two models were chosen and overviewed: MDRN and GFMN

**II. Chosen models inference:**

- GFMN was better on demo inference by PSNR and SSIM metrics

**III. Models training for additional tasks:**

- MDRN showed better results on additional tasks after fine-tuning with data augmentation

**IV. How to improve the perceptual quality of generated image by loss modifications**

- Several hypotheses have been proposed for quality and complexity improvement

https://github.com/RazinAleksandr/ISR_Task_2023

# Links


1. Li, Yawei, et al. "NTIRE 2023 challenge on efficient super-resolution: Methods and results." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
2. Mao, Yanyu, et al. "Multi-Level Dispersion Residual Network for Efficient Image Super-Resolution." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
3. Sun, Long, et al. "Spatially-Adaptive Feature Modulation for Efficient Image Super-Resolution." arXiv preprint arXiv:2302.13800 (2023).