Feature Distillation Interaction Weighting Network for Lightweight Image Super-Resolution

Guangwei Gao\textsuperscript{1†}, Wenjie Li\textsuperscript{1†}, Juncheng Li\textsuperscript{2*}, Fei Wu\textsuperscript{1}, Huimin Lu\textsuperscript{3}, Yi Yu\textsuperscript{4}

\textsuperscript{1}Nanjing University of Posts and Telecommunications \textsuperscript{2}The Chinese University of Hong Kong \textsuperscript{3}Kyushu Institute of Technology \textsuperscript{4}National Institute of Informatics

Reporter: Wenjie Li

https://github.com/IVIPLab/FDIWN
Outline

- Background & Related Works
- Motivation
- Feature Distillation Interaction Weighting Network
- Experiments & Discussion
- Summary
Outline

- Background & Related Works
- Motivation
- Feature Distillation Interaction Weighting Network
- Experiments & Discussion
- Summary
Background & Related Works

- Image Super-resolution
  - The purpose of single-image super-resolution (SISR) is to reconstruct a high-resolution (HR) image from its degraded low-resolution (LR) counterpart.

- Lightweight Image Super-resolution
  - Lightweight SR models are widely concerned for saving memory resources and computing resources.
  - In the case of fewer parameters and computation, a better performance is obtained.
How to reconstruct SR images?

Model

CARN/IDN/IMDN/MSFIN...

LR

DIV2K

SR

Background & Related Works
Background & Related Works

- **Ahn's CARN**


- **Hui's IDN**

Background & Related Works

- **Hui's**


- **Li's s-LWSR**


- **Wang's MSFIN**

Background & Related Works

- Other Lightweight Methods

**SRCNN**

**AWSRN**

**VDSR**

**RFDN**

**FSRCNN**

**LatticeNet**
Outline

- Background & Related Works
- Motivation
- Feature Distillation Interaction Weighting Network
- Experiments & Discussion
- Summary
In our proposed model, we believe that we should take advantage of both channel attention and spatial attention to obtain better SR performance.
Motivation

- Stronger feature extraction capabilities

Since WDSR has a broad activation mechanism, its information extraction ability will be better. In order for the model to focus on the important information, why not make the SA attention focus on each WDSR unit and adjust the output adaptively?
Motivation

■ **Target**
- We aim to explore a lightweight and efficient SISR model.

■ **Contributions**
- We propose a wide-residual attention weighting unit for lightweight SISR, including Wide Identical Residual Weighting (WIRW) unit and Wide Convolutional Residual Weighting (WCRW) unit, which has stronger feature distillation capabilities than ordinary residual blocks.
- We propose a novel Self-Calibration Fusion (SCF) module to replace the traditional concatenate operation for efficient feature interaction and fusion, which can aggregate more representative features and self-calibrate the input and output features.
- We propose a Wide-Residual Distillation Connection (WRDC) framework, which connects the coarse and distilled fine features within the module and allows features from different scales to interact with each other.
- We design a Feature Shuffle Weighted Group (FSWG) for pairwise feature fusion, which consists of a series of interactional WDIBs.
Outline

- Background & Related Works
- Motivation
- Feature Distillation Interaction Weighting Network
- Experiments & Discussion
- Summary
Network Architecture

Loss function: \( \hat{\theta} = \arg \min_\theta \frac{1}{N} \sum_{i=1}^{N} \| F_\theta(I_{LR}^i) - I_{HR}^i \|_1 \)
FDIWN-WDIB

- Wide-residual Distillation Interaction Block

- The $M_i$ and $N_i$ is the combination coefficient learning.
- The SA is the Shuffle Attention.
FDIWN-SA

- Shuffle Attention Mechanism

- F_{GP}: Global Pooling
- F_{GN}: Group Norm
- F_{C}: Fully Connected
- \otimes: Multiply Operation
- \sigma: Sigmoid
FDIWN

- **Combination Coefficient Learning**

\[
H_{GC}(x_c) = \sqrt{\frac{1}{HW} \sum_{(i,j) \in x_c} (x_c^{i,j} - \frac{1}{HW} \sum_{(i,j) \in x_c} x_c^{i,j})^2}
\]
Outline

- Background & Related Works
- Motivation
- Feature Distillation Interaction Weighting Network
- Experiments & Discussion
- Summary
Experiments & Discussion

- Quantitative comparisons

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Scale</th>
<th>Params</th>
<th>Multi-adds</th>
<th>Set5 PSNR</th>
<th>Set5 SSIM</th>
<th>Set14 PSNR</th>
<th>Set14 SSIM</th>
<th>BSDS100 PSNR</th>
<th>BSDS100 SSIM</th>
<th>Urban100 PSNR</th>
<th>Urban100 SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCNN (Dong et al. 2015)</td>
<td>57K</td>
<td>52.7G</td>
<td></td>
<td>32.75</td>
<td>0.9090</td>
<td>29.30</td>
<td>0.8215</td>
<td>28.41</td>
<td>0.7863</td>
<td>26.24</td>
<td>0.7889</td>
</tr>
<tr>
<td>FSRCNN (Dong, Loy, and Tang 2016)</td>
<td>12K</td>
<td>5.0G</td>
<td></td>
<td>33.16</td>
<td>0.9140</td>
<td>29.43</td>
<td>0.8242</td>
<td>28.53</td>
<td>0.7910</td>
<td>26.43</td>
<td>0.8080</td>
</tr>
<tr>
<td>VDSR (Kim, Lee, and Lee 2016a)</td>
<td>665K</td>
<td>612.6G</td>
<td></td>
<td>33.67</td>
<td>0.9210</td>
<td>29.78</td>
<td>0.8320</td>
<td>28.83</td>
<td>0.7990</td>
<td>27.14</td>
<td>0.8290</td>
</tr>
<tr>
<td>DRCN (Kim, Lee, and Lee 2016b)</td>
<td>1774K</td>
<td>1797.4G</td>
<td></td>
<td>33.82</td>
<td>0.9226</td>
<td>29.76</td>
<td>0.8311</td>
<td>28.80</td>
<td>0.7963</td>
<td>27.15</td>
<td>0.8276</td>
</tr>
<tr>
<td>IDN (Hui, Wang, and Gao 2018)</td>
<td>590K</td>
<td>105.6G</td>
<td></td>
<td>34.11</td>
<td>0.9253</td>
<td>29.99</td>
<td>0.8354</td>
<td>28.95</td>
<td>0.8013</td>
<td>27.42</td>
<td>0.8359</td>
</tr>
<tr>
<td>CARN-M (Ahn, Kang, and Sohn 2018)</td>
<td>412K</td>
<td>46.1G</td>
<td></td>
<td>33.99</td>
<td>0.9236</td>
<td>30.08</td>
<td>0.8367</td>
<td>28.91</td>
<td>0.8000</td>
<td>27.55</td>
<td>0.8385</td>
</tr>
<tr>
<td>CARN (Ahn, Kang, and Sohn 2018)</td>
<td>1592K</td>
<td>118.8G</td>
<td></td>
<td>34.29</td>
<td>0.9255</td>
<td>30.29</td>
<td>0.8407</td>
<td>29.06</td>
<td>0.8034</td>
<td>28.06</td>
<td>0.8493</td>
</tr>
<tr>
<td>IMDN (Hui et al. 2019)</td>
<td>703K</td>
<td>71.5G</td>
<td></td>
<td>34.36</td>
<td>0.9270</td>
<td>30.32</td>
<td>0.8417</td>
<td>29.09</td>
<td>0.8046</td>
<td>28.17</td>
<td>0.8519</td>
</tr>
<tr>
<td>AWSRN-M (Wang, Li, and Shi 2019)</td>
<td>1143K</td>
<td>116.6G</td>
<td></td>
<td>34.42</td>
<td>0.9275</td>
<td>30.32</td>
<td>0.8419</td>
<td>29.13</td>
<td>0.8059</td>
<td>28.26</td>
<td>0.8545</td>
</tr>
<tr>
<td>MADNet (Lan et al. 2020)</td>
<td>930K</td>
<td>88.4G</td>
<td></td>
<td>34.16</td>
<td>0.9253</td>
<td>30.21</td>
<td>0.8398</td>
<td>28.98</td>
<td>0.8023</td>
<td>27.77</td>
<td>0.8439</td>
</tr>
<tr>
<td>RFDN (Liu, Tang, and Wu 2020)</td>
<td>541K</td>
<td>55.4G</td>
<td></td>
<td>34.41</td>
<td>0.9273</td>
<td>30.34</td>
<td>0.8420</td>
<td>29.09</td>
<td>0.8050</td>
<td>28.21</td>
<td>0.8525</td>
</tr>
<tr>
<td>MAFFSRN (Muqet et al. 2020)</td>
<td>418K</td>
<td>34.2G</td>
<td></td>
<td>34.32</td>
<td>0.9269</td>
<td>30.35</td>
<td>0.8429</td>
<td>29.09</td>
<td>0.8052</td>
<td>28.13</td>
<td>0.8521</td>
</tr>
<tr>
<td>LAPAR-A (Li et al. 2021)</td>
<td>594K</td>
<td>114G</td>
<td></td>
<td>34.36</td>
<td>0.9267</td>
<td>30.34</td>
<td>0.8421</td>
<td>29.11</td>
<td>0.8054</td>
<td>28.15</td>
<td>0.8523</td>
</tr>
<tr>
<td>FDIWN-M (Ours)</td>
<td>446K</td>
<td>35.9G</td>
<td></td>
<td>34.46</td>
<td>0.9274</td>
<td>30.35</td>
<td>0.8423</td>
<td>29.10</td>
<td>0.8051</td>
<td>28.16</td>
<td>0.8528</td>
</tr>
<tr>
<td>FDIWN (Ours)</td>
<td>645K</td>
<td>51.5G</td>
<td></td>
<td>34.52</td>
<td>0.9281</td>
<td>30.42</td>
<td>0.8438</td>
<td>29.14</td>
<td>0.8065</td>
<td>28.36</td>
<td>0.8567</td>
</tr>
</tbody>
</table>

- Visual results of FDIWN with other SR methods (x3)
Experiments & Discussion

- Visual results of FDIWN with other SR methods (x3)
## Experiments & Discussion

- **Quantitative comparisons**

<table>
<thead>
<tr>
<th>Method</th>
<th>LR Size</th>
<th>PSNR</th>
<th>SSIM</th>
<th>MS-SSIM</th>
<th>PSNR</th>
<th>SSIM</th>
<th>MS-SSIM</th>
<th>PSNR</th>
<th>SSIM</th>
<th>MS-SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCNN (Dong et al. 2015)</td>
<td>57K</td>
<td>52.7G</td>
<td>30.48</td>
<td>0.8628</td>
<td>27.49</td>
<td>0.7503</td>
<td>26.90</td>
<td>0.7101</td>
<td>24.52</td>
<td>0.7221</td>
</tr>
<tr>
<td>FSRCNN (Dong, Loy, and Tang 2016)</td>
<td>12K</td>
<td>4.6G</td>
<td>30.71</td>
<td>0.8657</td>
<td>27.59</td>
<td>0.7535</td>
<td>26.98</td>
<td>0.7150</td>
<td>24.62</td>
<td>0.7280</td>
</tr>
<tr>
<td>VDSR (Kim, Lee, and Lee 2016a)</td>
<td>635K</td>
<td>612.6G</td>
<td>31.35</td>
<td>0.8838</td>
<td>28.01</td>
<td>0.7674</td>
<td>27.29</td>
<td>0.7251</td>
<td>25.18</td>
<td>0.7524</td>
</tr>
<tr>
<td>DRCN (Kim, Lee, and Lee 2016b)</td>
<td>1774K</td>
<td>1797.4G</td>
<td>31.53</td>
<td>0.8854</td>
<td>28.02</td>
<td>0.7670</td>
<td>27.23</td>
<td>0.7233</td>
<td>25.14</td>
<td>0.7510</td>
</tr>
<tr>
<td>LapSRN (Lai et al. 2017)</td>
<td>813K</td>
<td>149.4G</td>
<td>31.54</td>
<td>0.8850</td>
<td>28.19</td>
<td>0.7720</td>
<td>27.32</td>
<td>0.7280</td>
<td>25.21</td>
<td>0.7560</td>
</tr>
<tr>
<td>IDN (Hui, Wang, and Gao 2018)</td>
<td>590K</td>
<td>81.9G</td>
<td>31.82</td>
<td>0.8903</td>
<td>28.25</td>
<td>0.7730</td>
<td>27.41</td>
<td>0.7297</td>
<td>25.41</td>
<td>0.7632</td>
</tr>
<tr>
<td>CARN-M (Ahn, Kang, and Sohn 2018)</td>
<td>412K</td>
<td>32.5G</td>
<td>31.92</td>
<td>0.8903</td>
<td>28.42</td>
<td>0.7762</td>
<td>27.44</td>
<td>0.7304</td>
<td>25.62</td>
<td>0.7694</td>
</tr>
<tr>
<td>CARN (Ahn, Kang, and Sohn 2018)</td>
<td>1592K</td>
<td>90.9G</td>
<td>32.13</td>
<td>0.8937</td>
<td>28.60</td>
<td>0.7806</td>
<td>27.58</td>
<td>0.7349</td>
<td>26.07</td>
<td>0.7837</td>
</tr>
<tr>
<td>IMDN (Hui et al. 2019)</td>
<td>715K</td>
<td>40.9G</td>
<td>32.21</td>
<td>0.8948</td>
<td>28.58</td>
<td>0.7811</td>
<td>27.56</td>
<td>0.7353</td>
<td>26.04</td>
<td>0.7838</td>
</tr>
<tr>
<td>AWSRN-M (Wang, Li, and Shi 2019)</td>
<td>1254K</td>
<td>72.0G</td>
<td>32.21</td>
<td><strong>0.8954</strong></td>
<td>28.65</td>
<td><strong>0.7832</strong></td>
<td>27.60</td>
<td>0.7368</td>
<td>26.15</td>
<td>0.7884</td>
</tr>
<tr>
<td>MADNet (Lan et al. 2020)</td>
<td>1002K</td>
<td>54.1G</td>
<td>31.95</td>
<td>0.8917</td>
<td>28.44</td>
<td>0.7780</td>
<td>27.47</td>
<td>0.7327</td>
<td>25.76</td>
<td>0.7746</td>
</tr>
<tr>
<td>RFDN (Liu, Tang, and Wu 2020)</td>
<td>550K</td>
<td>31.6G</td>
<td><strong>32.24</strong></td>
<td>0.8952</td>
<td>28.61</td>
<td>0.7819</td>
<td>27.57</td>
<td>0.7360</td>
<td>26.11</td>
<td>0.7858</td>
</tr>
<tr>
<td>AFFSRN (Muqueet et al. 2020)</td>
<td>441K</td>
<td>19.3G</td>
<td>32.18</td>
<td>0.8948</td>
<td>28.58</td>
<td>0.7812</td>
<td>27.57</td>
<td>0.7361</td>
<td>26.04</td>
<td>0.7848</td>
</tr>
<tr>
<td>ECBSR (Zhang, Zeng, and Zhang 2021)</td>
<td>603K</td>
<td>34.7G</td>
<td>31.92</td>
<td>0.8946</td>
<td>28.34</td>
<td>0.7817</td>
<td>27.48</td>
<td>0.7393</td>
<td>25.81</td>
<td>0.7773</td>
</tr>
<tr>
<td>LAPAR-A (Li et al. 2021)</td>
<td>659K</td>
<td>94G</td>
<td>32.15</td>
<td>0.8944</td>
<td>28.61</td>
<td>0.7818</td>
<td><strong>27.61</strong></td>
<td>0.7366</td>
<td>26.14</td>
<td>0.7871</td>
</tr>
<tr>
<td>FDIWN-M (Ours)</td>
<td>454K</td>
<td>19.6G</td>
<td>32.17</td>
<td>0.8941</td>
<td>28.55</td>
<td>0.7806</td>
<td>27.58</td>
<td>0.7364</td>
<td>26.02</td>
<td>0.7844</td>
</tr>
<tr>
<td>FDIWN (Ours)</td>
<td>664K</td>
<td>28.4G</td>
<td><strong>32.23</strong></td>
<td><strong>0.8955</strong></td>
<td><strong>28.66</strong></td>
<td><strong>0.7829</strong></td>
<td><strong>27.62</strong></td>
<td><strong>0.7380</strong></td>
<td><strong>26.28</strong></td>
<td><strong>0.7919</strong></td>
</tr>
</tbody>
</table>

- **Visual results of FDIWN with other SR methods (x4)**

![Visual results](image.png)
Experiments & Discussion

- Visual results of FDIWN with other SR methods (x4)
Experiments & Discussion

- **Efficiency trade-off**

- **Model complexity analysis**
Experiments & Discussion

- Investigations of the model size and performance

- Feature visualization of different module
# Experiments & Discussion

- The combination structure of WDIB

<table>
<thead>
<tr>
<th>Method</th>
<th>BI</th>
<th>WIRW</th>
<th>Params</th>
<th>Multi-adds</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>✗</td>
<td>✗</td>
<td>215K</td>
<td>22.0G</td>
<td>37.81</td>
<td>0.9598</td>
</tr>
<tr>
<td>FDIWN</td>
<td>✓</td>
<td>✗</td>
<td>225K</td>
<td>24.4G</td>
<td>37.85</td>
<td><strong>0.9600</strong></td>
</tr>
<tr>
<td>FDIWN</td>
<td>✓</td>
<td>✓</td>
<td>230K</td>
<td>24.4G</td>
<td><strong>37.88</strong></td>
<td><strong>0.9600</strong></td>
</tr>
</tbody>
</table>

- The effectiveness of WIRW and WCRW

<table>
<thead>
<tr>
<th>Case</th>
<th>Method</th>
<th>Channels</th>
<th>Params</th>
<th>Multi-adds</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baseline</td>
<td>24</td>
<td>152K</td>
<td>23.2G</td>
<td>37.70</td>
<td>0.9594</td>
</tr>
<tr>
<td>2</td>
<td>FDIWN</td>
<td>48</td>
<td>96K</td>
<td>9.7G</td>
<td>37.64</td>
<td>0.9592</td>
</tr>
<tr>
<td>3</td>
<td>FDIWN</td>
<td>120</td>
<td>131K</td>
<td>9.7G</td>
<td><strong>37.72</strong></td>
<td><strong>0.9596</strong></td>
</tr>
</tbody>
</table>

- The effectiveness of WRDC and SCF

<table>
<thead>
<tr>
<th>Method</th>
<th>WR</th>
<th>DC</th>
<th>SCF</th>
<th>Params</th>
<th>Multi-adds</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline1</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>59K</td>
<td>3.3G</td>
<td>37.52</td>
<td>0.9587</td>
</tr>
<tr>
<td>Baseline2</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>59K</td>
<td>4.9G</td>
<td>37.53</td>
<td>0.9589</td>
</tr>
<tr>
<td>FDIWN</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>89K</td>
<td>6.5G</td>
<td>37.58</td>
<td>0.9591</td>
</tr>
<tr>
<td>FDIWN</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>65K</td>
<td>6.5G</td>
<td>37.59</td>
<td>0.9590</td>
</tr>
<tr>
<td>FDIWN</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>96K</td>
<td>9.7G</td>
<td><strong>37.64</strong></td>
<td><strong>0.9592</strong></td>
</tr>
</tbody>
</table>
Outline

- Background & Related Works
- Motivation
- Feature Distillation Interaction Weighting Network
- Experiments & Discussion
- Summary
The specially designed wide-residual weighting units (including WIRW and WCRW) have a stronger ability to distill useful features than ordinary residual blocks.

The well designed wide-residual units based WRDC module and SCF module can flexibly aggregate and distill more representative features.

The experiment show that our proposed FDIWN achieved a good balance between model size, performance, and computational cost.
Thanks !