DLSS 4 - Super Resolution

A Transformer-based Model

14 - Jiahao Song

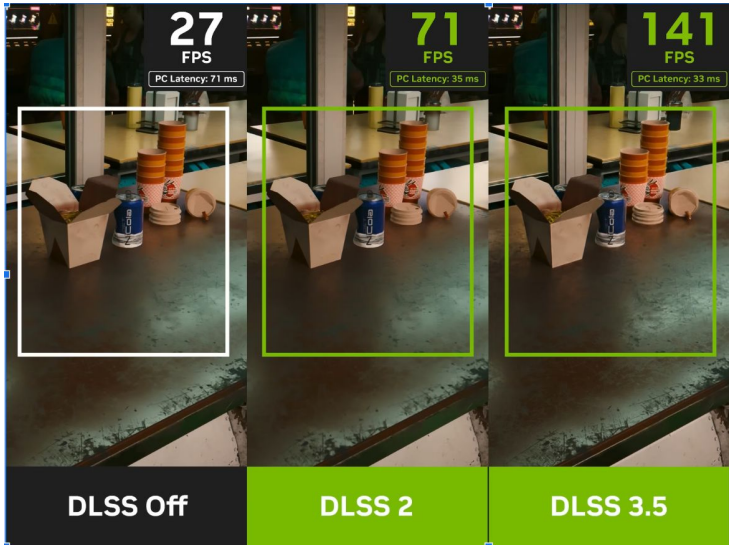
NOTE: DLSS 4 paper **has not been released yet,** therefore, the theory part will use DRCT as a substitution. (implementation not included)

DLSS

- Developed by **NVIDIA**
- A suite of neural **rendering** technologies

DLSS

- Developed by **NVIDIA**
- A suite of neural **rendering** technologies
- **Boosts frames rates (Frame Generation)**



DLSS

- Developed by **NVIDIA**
- A suite of neural **rendering** technologies
- Boosts frames rates (Frame Generation)
- **Delivers crisp, high-quality images (CNN-based Super Resolution)**

DLSS 4 (SOTA)

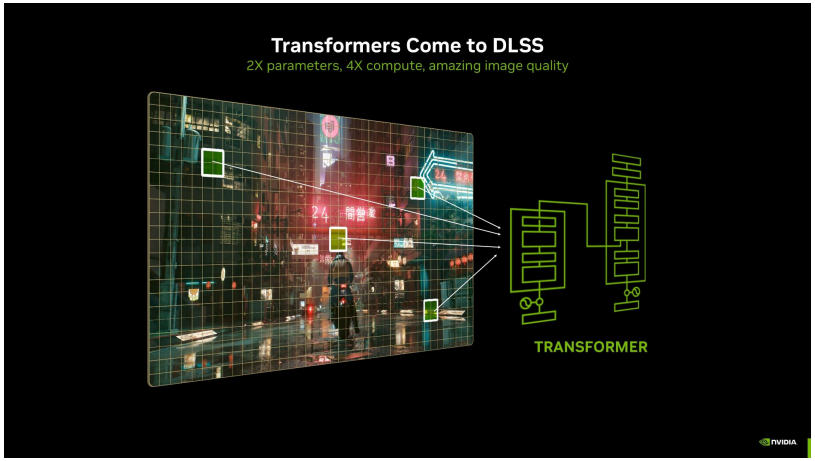
- Multi Frame Generation (2x, 3x, 4x)

DLSS 4 (SOTA)

- Multi Frame Generation (2x, 3x, 4x)
  - Boosts frames **significantly**, but **introduces input latency**

DLSS 4 (SOTA)

- Super Resolution Enhancement
  - **Transformer-based**



DLSS 4 (SOTA)

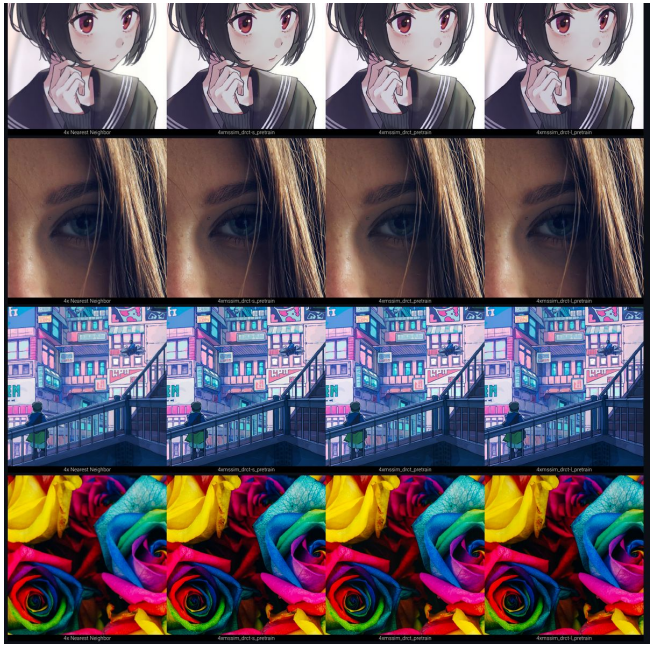
- Super Resolution Enhancement
  - Transformer-based
  - **Higher detail in motion, less ghosting and artifacts**

DLSS 4 (SOTA)

- Super Resolution Enhancement
  - Unfortunately, the DLSS 4 paper **is not released yet**, so I'll focus on a more general one called [DRCT]

DRCT

- Dense-Residual-Connected Transformer (DRCT)

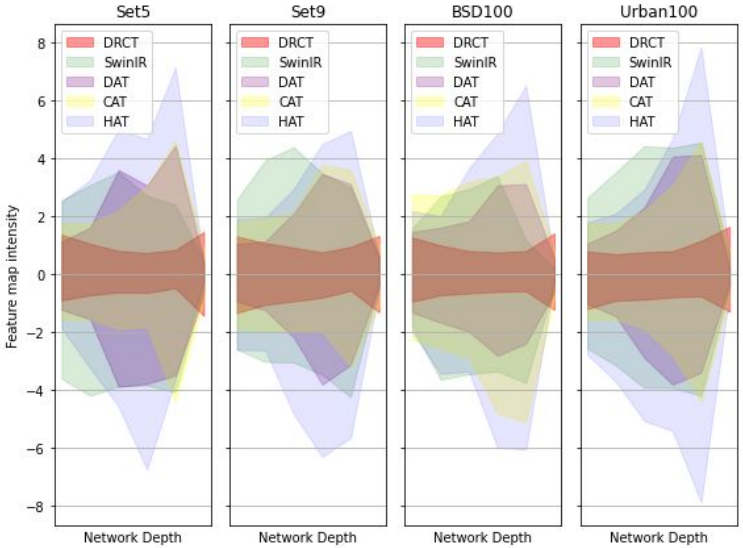


From left to right:
Nearest Neighbor - DRCT-S - DRCT - DRCT-L

[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

DRCT

- Dense-Residual-Connected Transformer (DRCT)
- **Solve the problem of information bottleneck in Swin-Transformer**
  - Swin-Transformer: window-sliding vision transformer

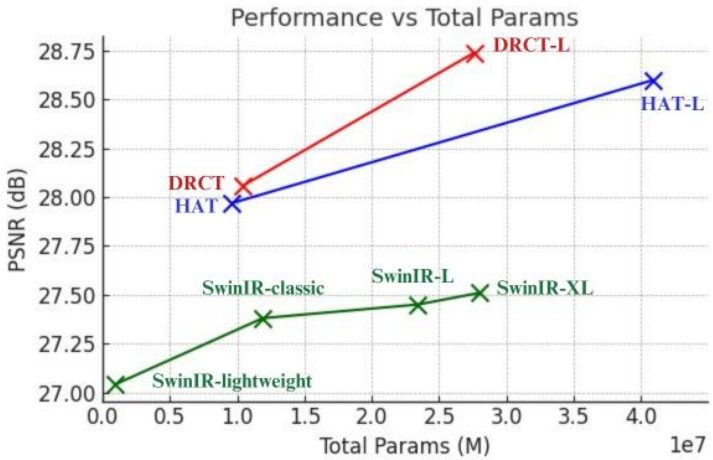


* DRCT maintains **stable feature map intensity** as the network becomes deeper, while the intensity of other transformer-based vision networks approaches to zero (losing “long-term memory”)

[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

DRCT

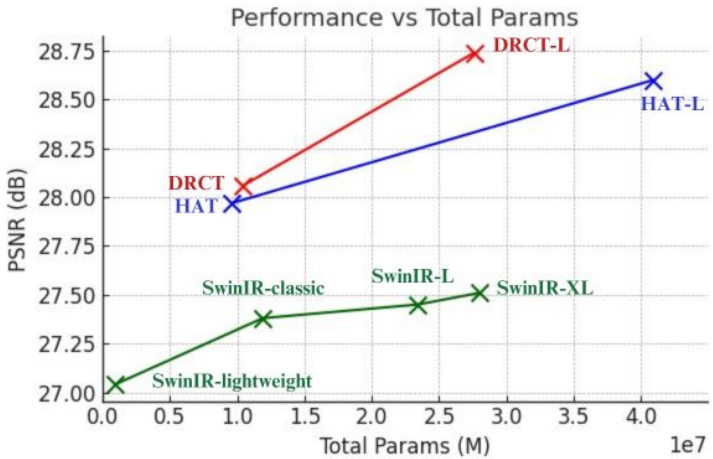
- Dense-Residual-Connected Transformer (DRCT)
- Solve the problem of information bottleneck in Swin-Transformer
- **Less parameters to get similar performance as Swin-Transformer**



[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

DRCT

- Dense-Residual-Connected Transformer (DRCT)
- Solve the problem of information bottleneck in Swin-Transformer
- **Less parameters to get similar performance as Swin-Transformer**



DRCT-L uses **nearly half of parameters** of HAT-L but gets better results

[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

Model Structure of DRCT

- Shallow Feature Extraction
- Deep Feature Extraction
- Image Reconstruction

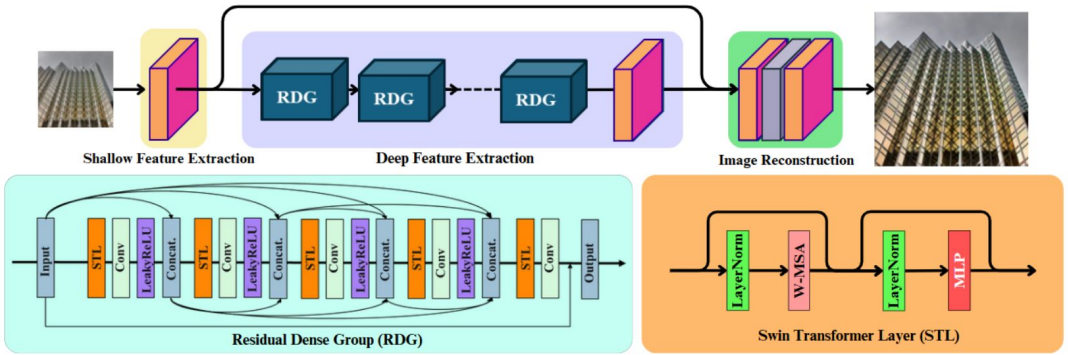


Figure 3. The overall architecture of the proposed Dense-residual-connected Transformer (DRCT) and the structure of Residual-Dense Group (RDG). Each RDG contains five consecutive Swin-Dense-Residual-Connected Blocks (SDRCBs). By integrating dense-connection [15] into SwinIR [34], the efficiency can be improved for *Saving Image Super-resolution away from Information Bottleneck.*

[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

Model Structure of DRCT

- **Shallow Feature Extraction**

$$F_0 = Conv(I_{LQ})$$

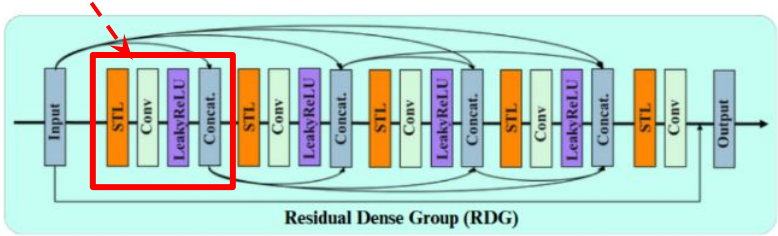
- The equation takes in a **low-quality image** into a 3x3 conv.
- Shallow features are the output of conv

[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

Model Structure of DRCT

- Shallow Feature Extraction
- **Deep Feature Extraction**

- Residual Dense Group (RDG)
  - Swin-Dense-Residual-Connected Block (SDRCB)
  - Captures **long-range dependency**



[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

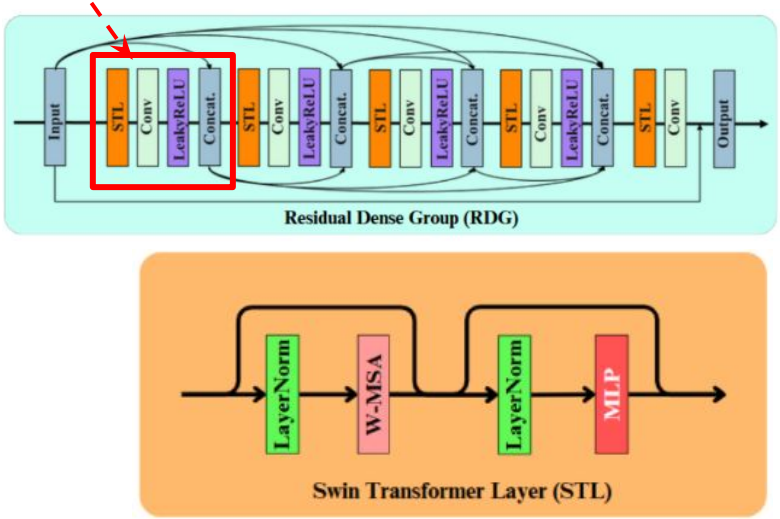
[2] https://sh-tsang.medium.com/review-swin-transformer-3438ea335585

Model Structure of DRCT

- Shallow Feature Extraction
- **Deep Feature Extraction**

- Residual Dense Group (RDG)
  - Swin-Dense-Residual-Connected Block (SDRCB)
  - Captures **long-range dependency**

- Swin Transformer Layer (STL)
  - Focuses on the **global** content of feature maps



[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

[2] https://sh-tsang.medium.com/review-swin-transformer-3438ea335585

Model Structure of DRCT

- Shallow Feature Extraction
- Deep Feature Extraction
- **Image Reconstruction**

$$I_{SR} = H_{rec}\left(F_0 + F_{DF} \right)$$

- Adds shallow and deep features
- Feeds it into the reconstruction layer
- Overall, low-quality images are **upscaled** to high-quality ones

[1] Hsu, Chih-Chung, Chia-Ming Lee, and Yi-Shiuan Chou. "DRCT: Saving Image Super-resolution away from Information Bottleneck." *arXiv preprint arXiv:2404.00722* (2024).

Thank you!