

cvpaper.challenge

# Image Super-Resolution Models

---

## 画像超解像モデル

大谷豪, 松尾雄斗, 児玉憲武

<http://xpaperchallenge.org/cv>

# 本資料の構成

---

- **イントロダクション**
  - 画像超解像の概要を説明
    - PSNR-oriented
    - Perception-oriented
- **画像超解像 (SR) モデル技術マップ**
- **論文サマリ**
  - 技術マップに載せたモデルの論文について簡単に紹介
- **メタサーベイ&最新トレンド**
  - 論文サーベイ&メンバー間の議論から見えたメタ要素
    - これまでの知見からわかること
    - これまでの知見でこれからの推移として予想できるもの

# 画像超解像とは？

画像の解像度を上げて鮮明にするアルゴリズム  
ほとんどの場合，DNNが使用される

128 × 128 px

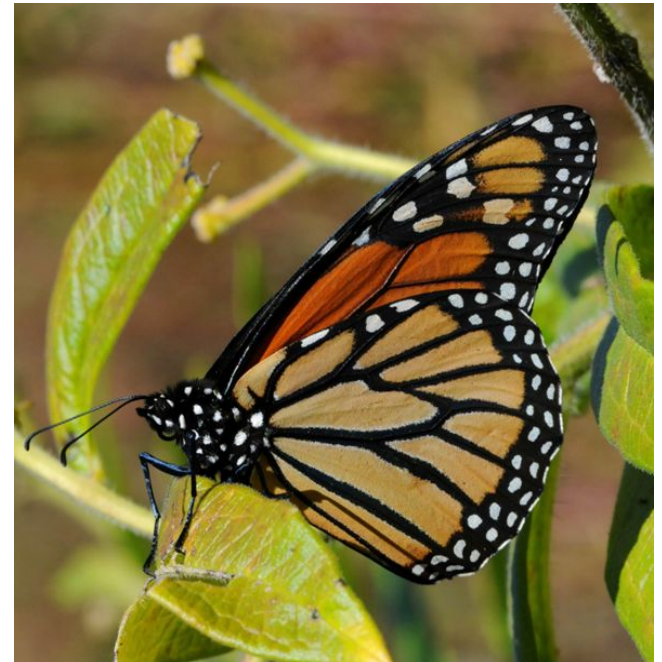


低解像度画像

256 × 256 px



2 ×  
upsampling

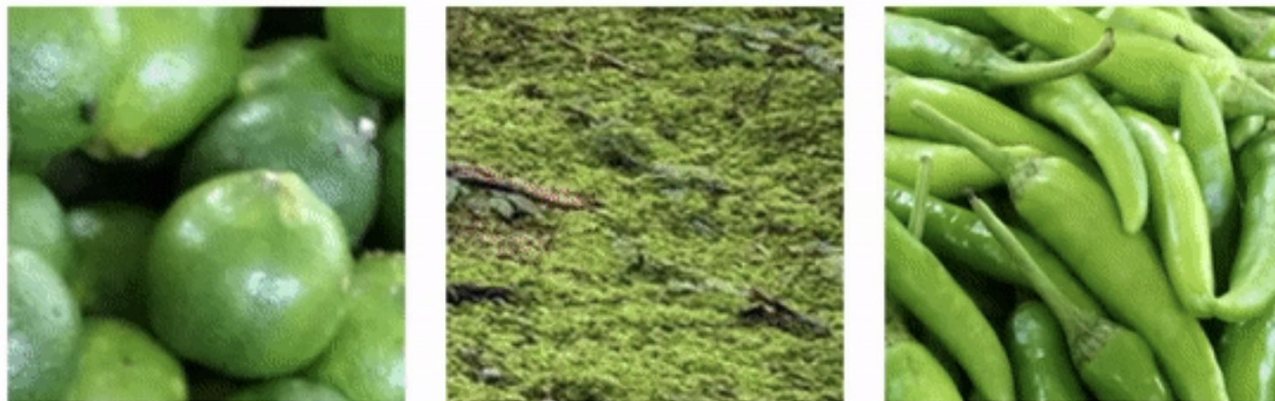


高解像度画像

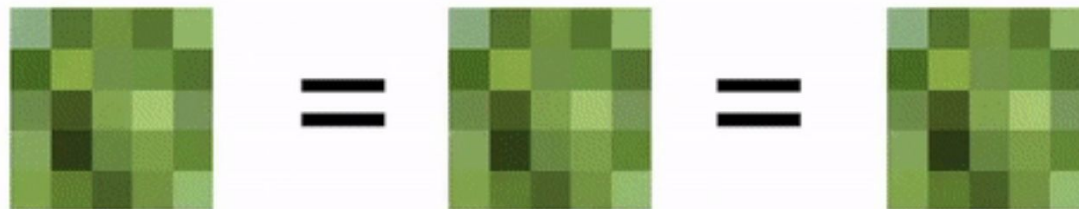
# ill-posed problem

低解像度画像からの復元解が複数存在

高解像度画像



低解像度画像





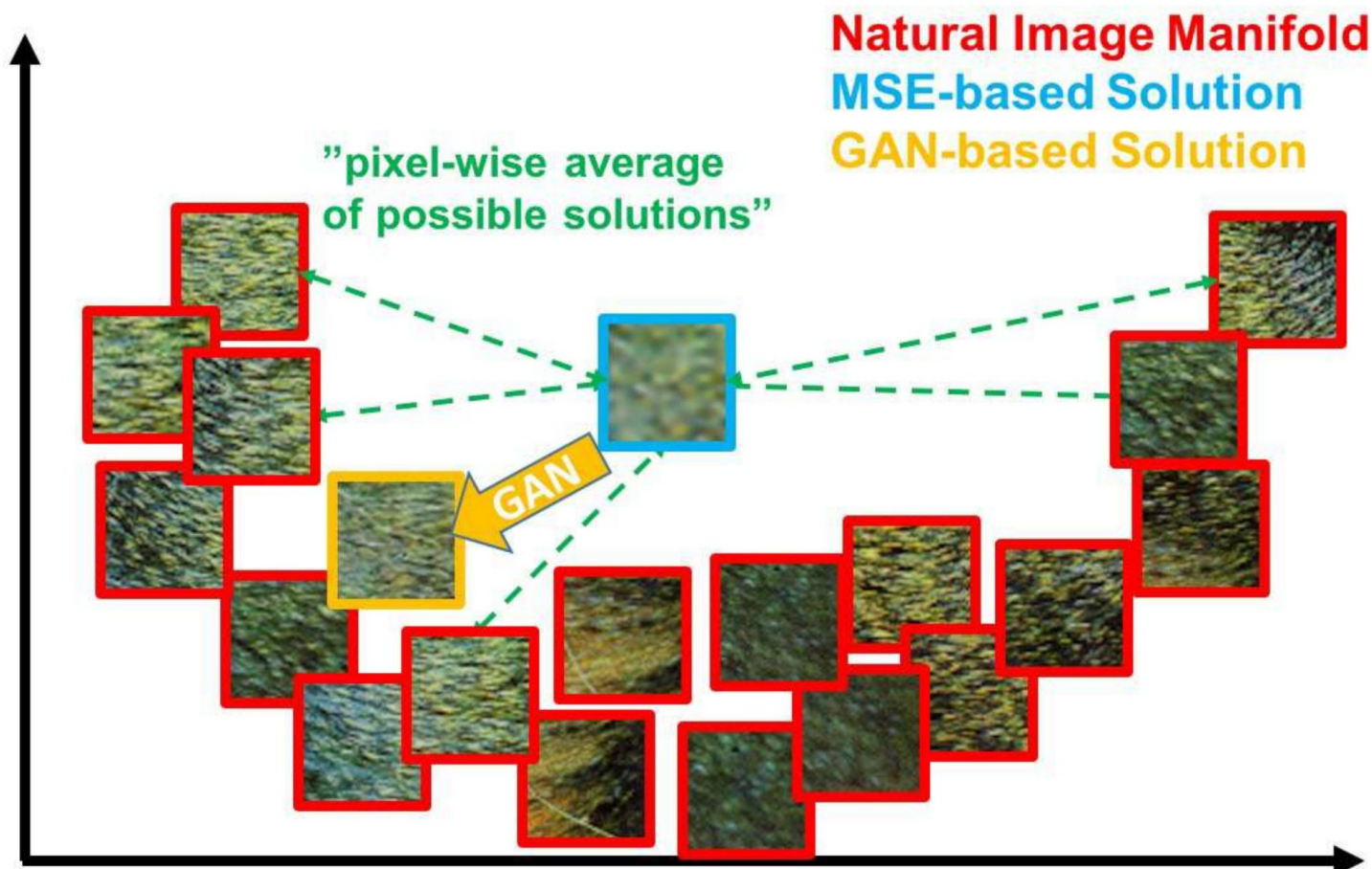
# 超解像手法の種類

---

- **PSNR-oriented (CNN, Transformer)**
  - ピクセル単位での整合性を重視
  - ボケが残存しやすい
- **Perception-oriented (GAN, Diffusion)**
  - 視覚的観点からの画像品質を重視
  - 生成モデルを使用するため実在しないものを作り出してしまふ可能性がある

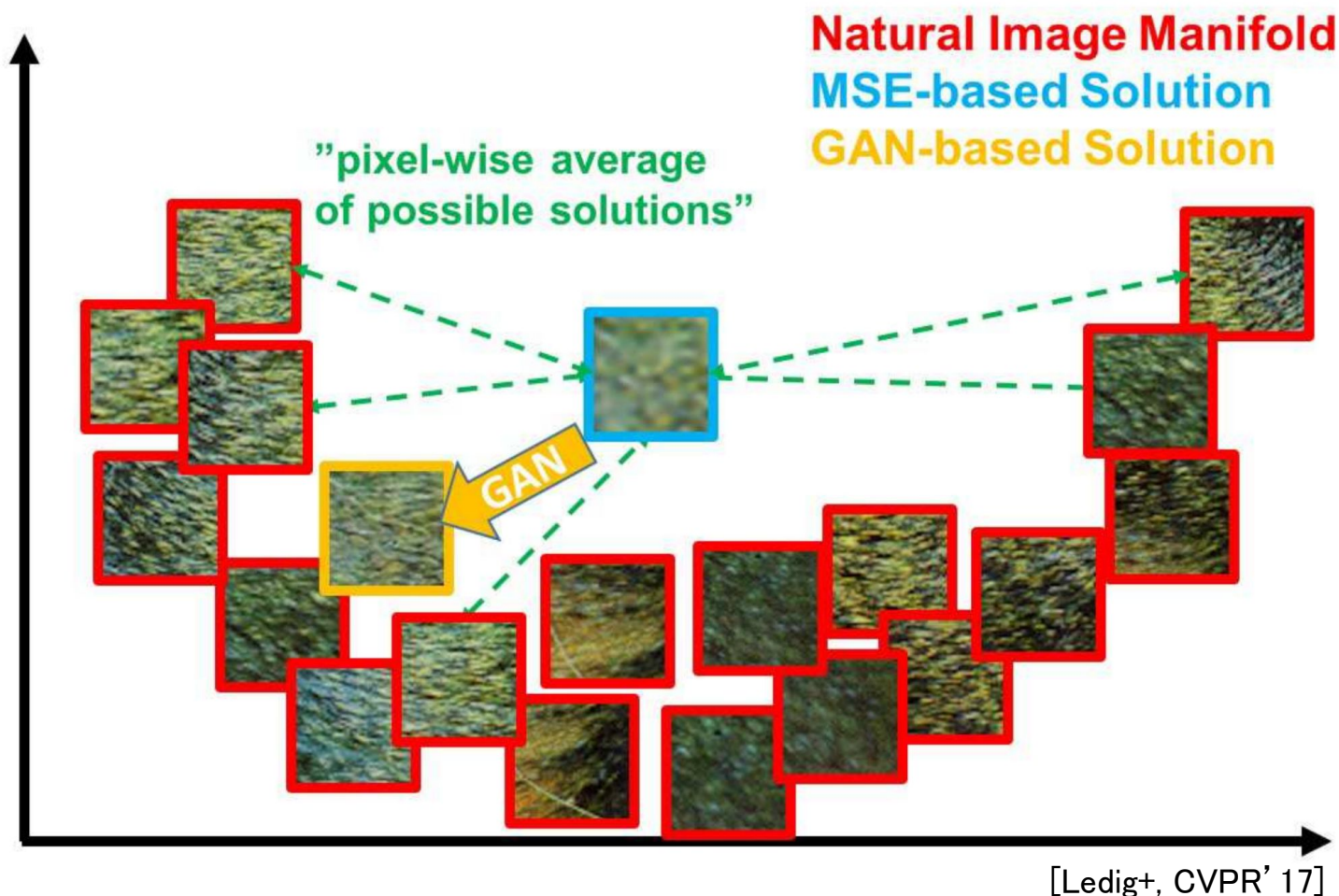
# PSNR-oriented vs Perception-oriented

PSNR-orientedな手法はGTとのピクセル単位での乖離は少ないがボケやすい



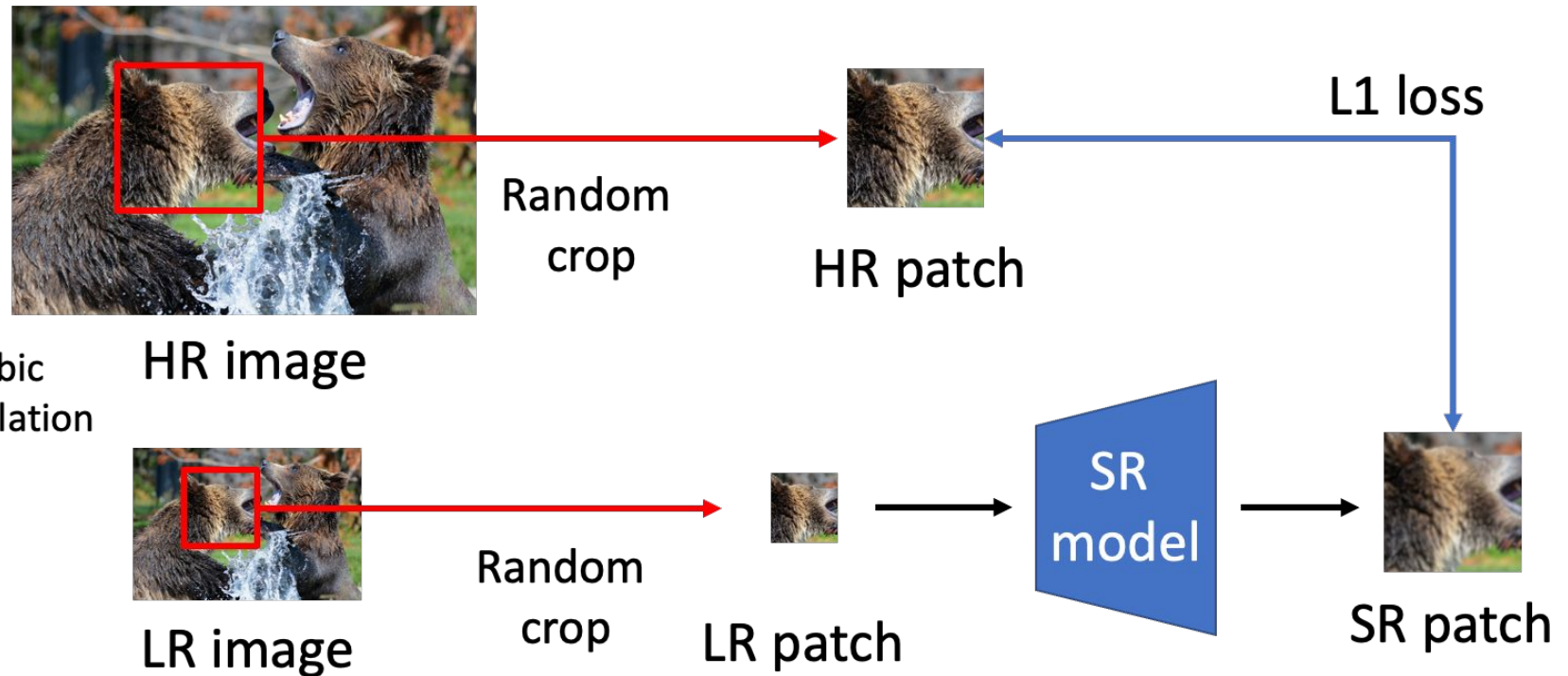
# PSNR-oriented vs Perception-oriented

Perception-orientedな手法はボケにくいですが GTとの  
ピクセル単位での乖離が大きくなりやすい



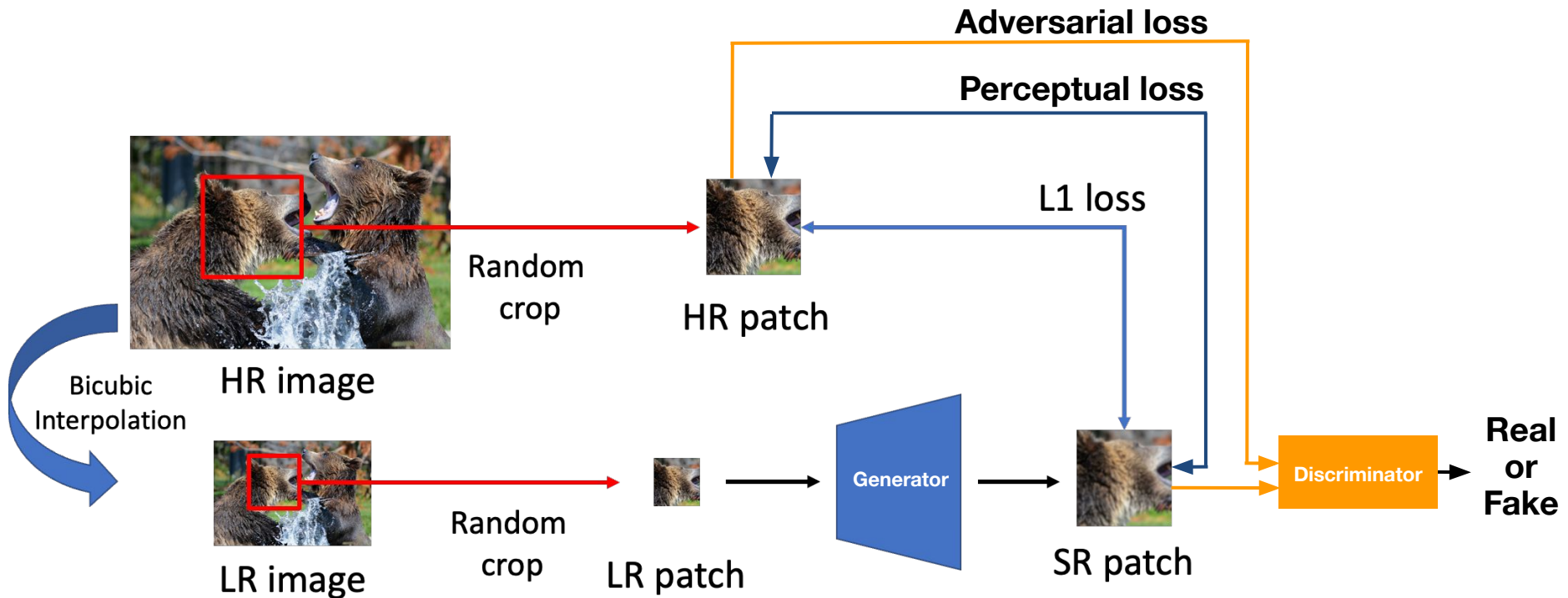
# 学習の流れ (PSNR-oriented)

損失関数は L1 loss のみのシンプルな構成



# 学習の流れ ( Perceptual-oriented, GAN )

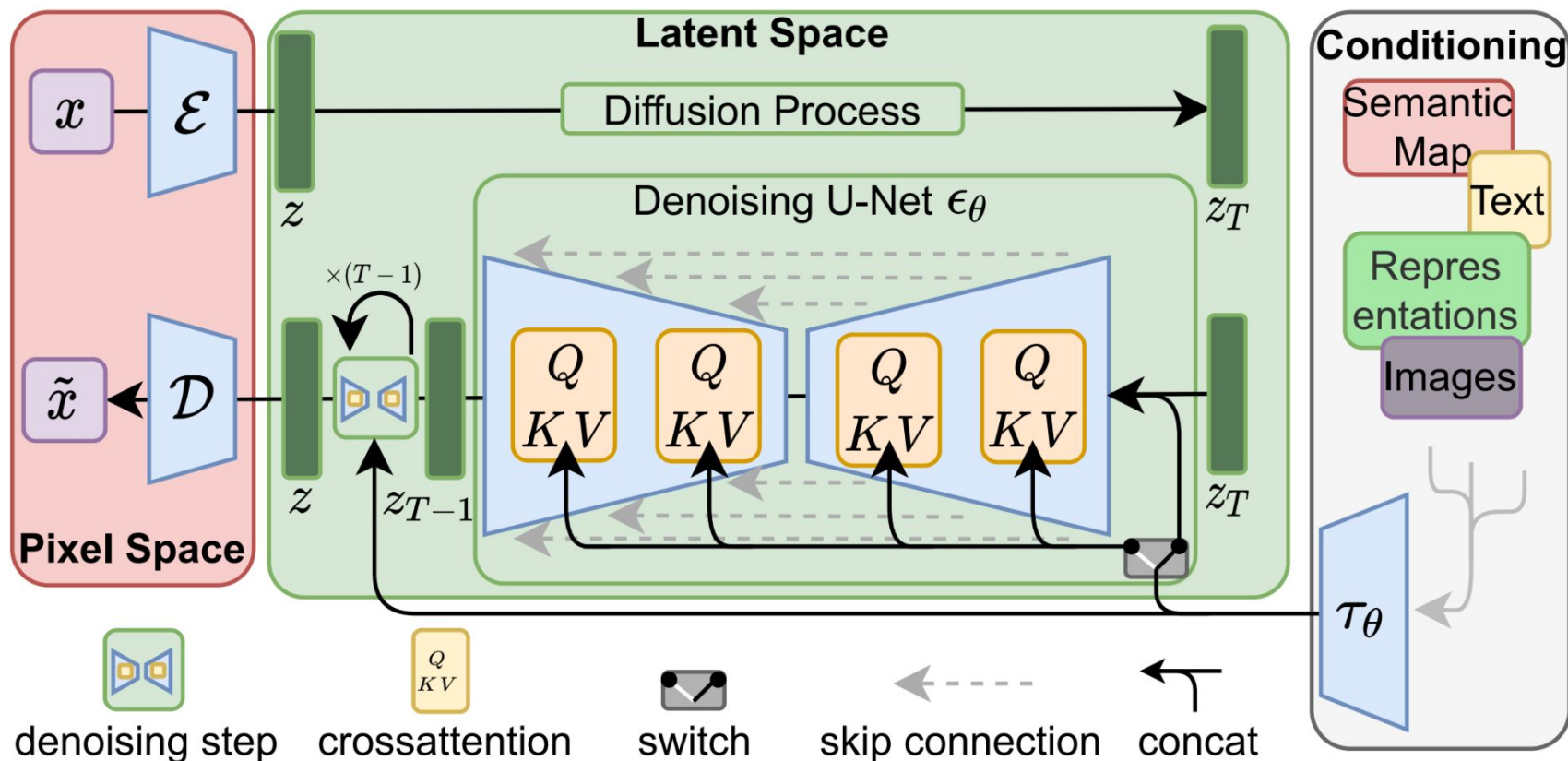
損失関数は L1 loss, perceptual loss, adversarial lossの構成が多い





# 学習の流れ (Perceptual-oriented, Diffusion)

近年はLDMベースの手法が多い  
条件付けにLR画像を使用



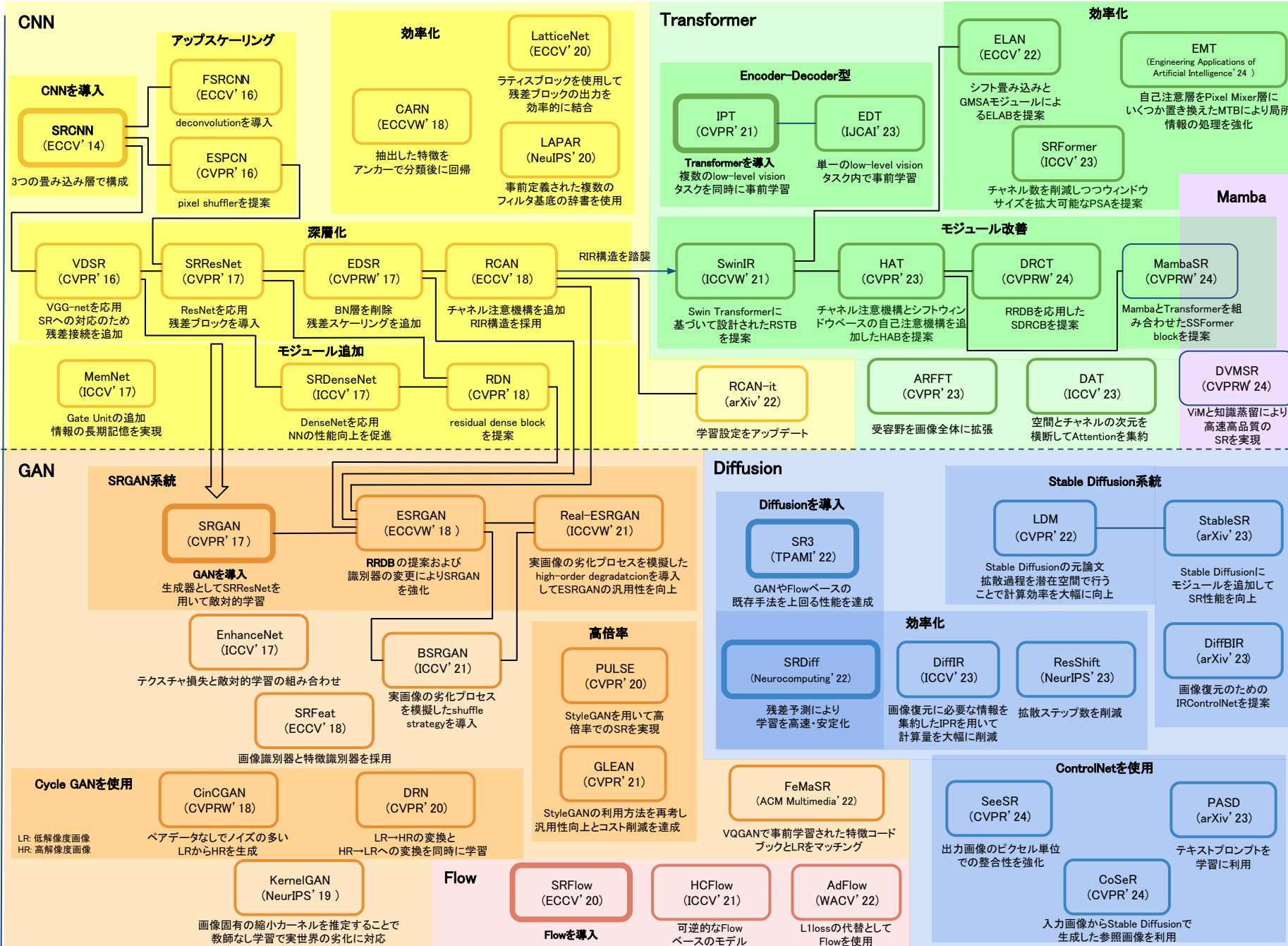
[Rombach+, CVPR' 22]

# 画像超解像 (SR) モデル技術マップ

大谷 豪 (慶應義塾大学), 児玉 憲武 (横浜国立大学), 松尾 雄斗 (東京電機大学)

PSNR単位での整合性を重視  
ピクセル単位での整合性を重視

Perception-oriented  
視覚的観点からの画像品質を重視





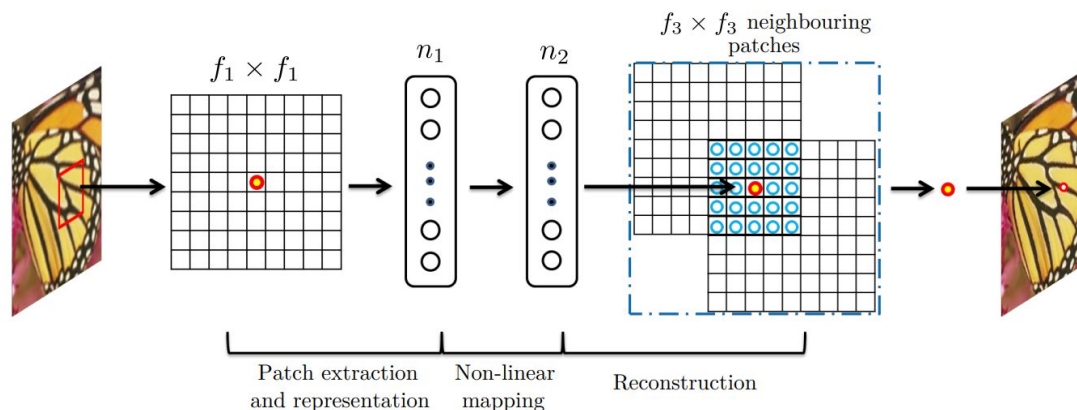
# 論文紹介

会議 : ECCV 2014

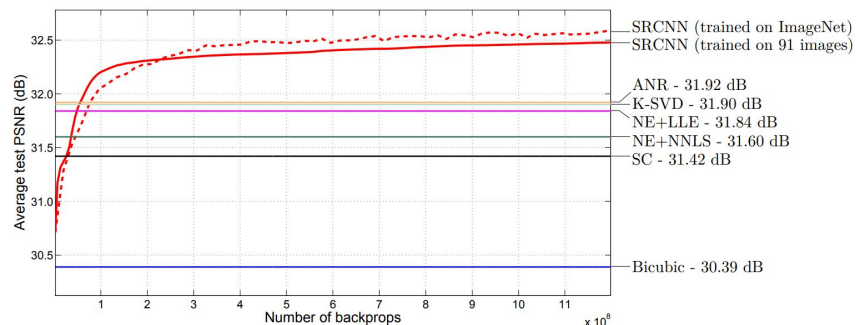
著者 : Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang

## SRCNN: CNNをSuper-Resolution (SR)に適用して当時のSOTAを達成

- 高速かつ高パフォーマンスなEnd2EndのSRをCNNで初めて実現



- モデルのトレーニングに大規模なデータセット (ImageNet)を用いることの有効性を示した
  - より大規模で長時間の訓練を行うことでパフォーマンスが上がることを示唆

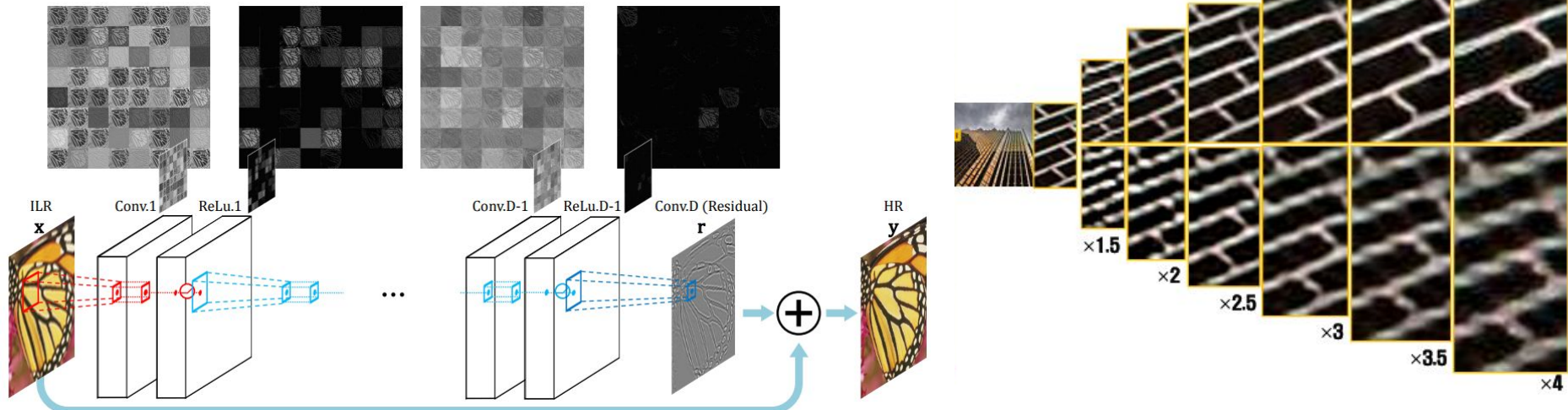


会議 : CVPR 2016

著者 : Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee

## VDSR: 20層の深いCNNをSRに導入して速度・精度ともにSOTAを達成

- 収束速度が遅くなるため困難であった深い層を持つモデルの訓練を実現
  - residual-learningと高いlearning rateを用いることで高速でモデルを最適化
  - 学習の安定性のために勾配クリッピングを導入
- residual-learningの導入により, 速度・精度ともにSOTAを達成
  - 何層にもわたって畳み込むことで画像のコンテキストを大域にわたって抽出可能
- 単一のモデルで複数のスケールを扱い, 単一のスケールを扱ったモデルと同等またはそれ以上の精度を達成



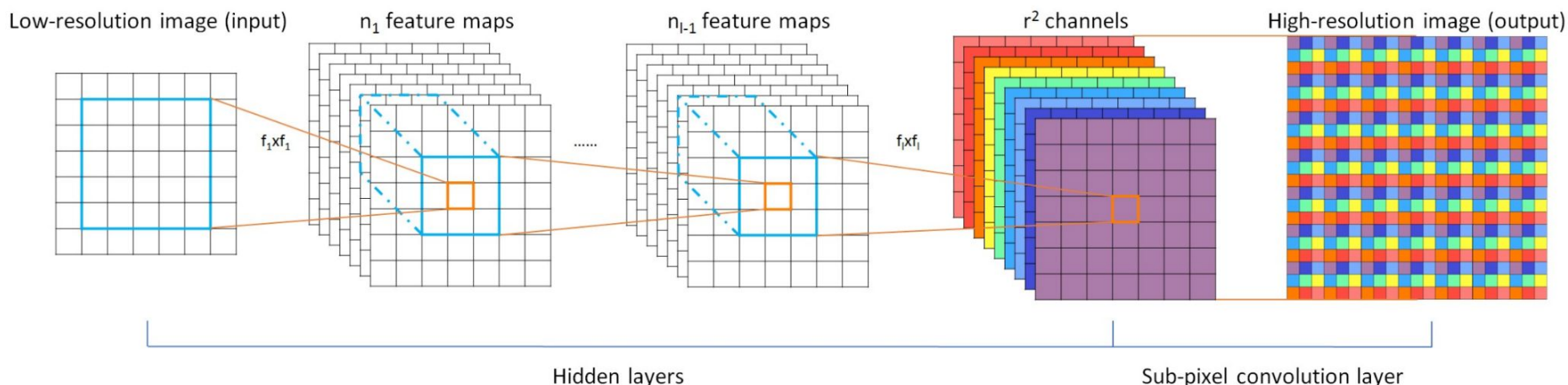
# Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network

会議 : CVPR 2016

著者 : Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan Wang

## ESPCN: 低計算量のupscale filterにより1080p videoに対応可能なSRを実現

- LRを入力としてLR空間で演算を行うことによりパラメータ数・計算量を削減
- LRからHRへupscaleするために学習可能なSub-pixel convolution layerを導入
  - ネットワーク最終層でupscaleを行う
  - 畳み込みによって複数のチャンネルを生成し, 再成型することでupscaleを行う
  - $H \times W \times C \cdot r^2$ のtensorから  $rH \times rW \times C$ に変換するためにshuffle演算を利用

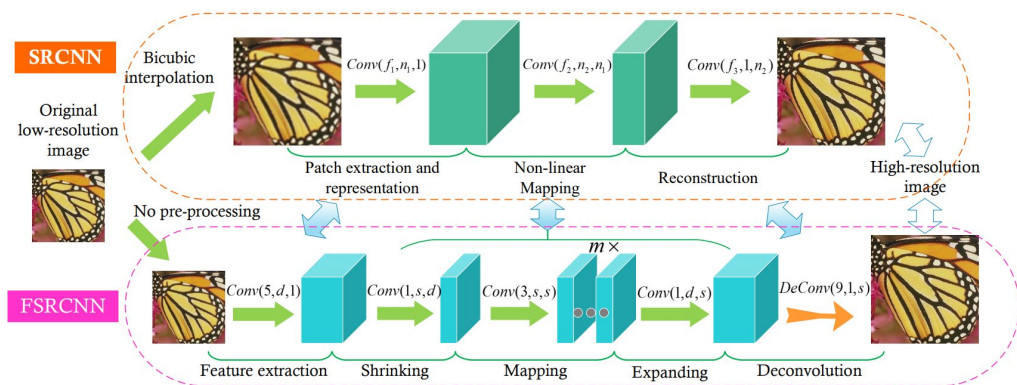


会議 : ECCV 2016

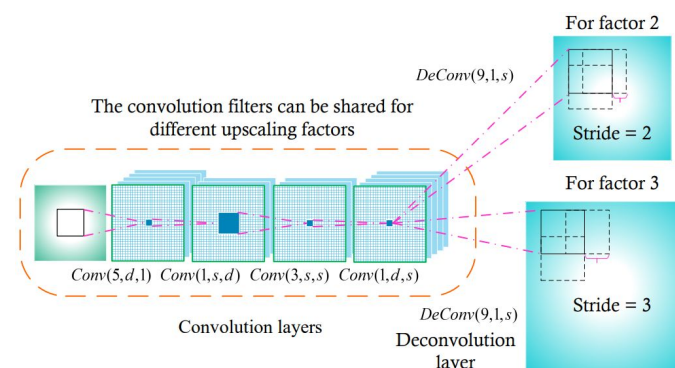
著者 : Chao Dong, Chen Change Loy, and Xiaoou Tang

## FSRCNN : SRCNNの構造を見直すことで40倍高速化

- 最終層にdeconvolution layerを導入し, LRからHRへのマッピングを直接学習
  - 畳み込みの逆演算とみなすことができ, スライドとupscaling factorが対応
  - deconvolution layerのみfinetuningすることで, 異なるupscaling factorに対して転用可能
- 特徴次元を圧縮してからマッピングし, その後拡張することでパラメータ数を削減
  - フィルターサイズ1の畳み込み層により次元を変更
  - より小さなフィルターサイズを採用するが, より多くのマッピング層を採用
- 活性化関数にPReLUを使用し, モデルのパラメータを最大限活用



SRCNN と FSRCNNの比較



deconvolution layer

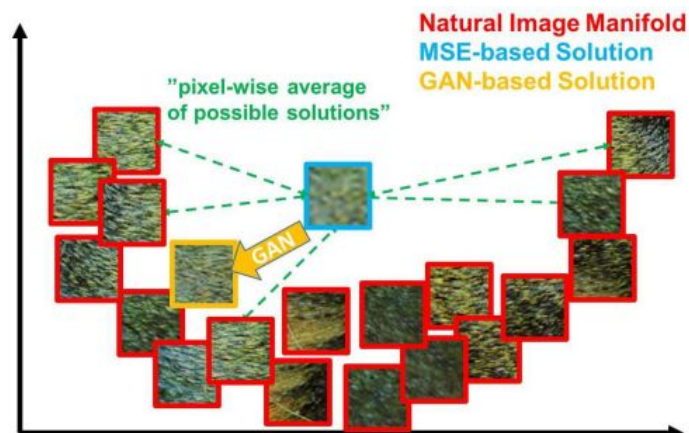


会議 : CVPR 2017

著者 : Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang and Wenzhe Shi

## SRResNet / SRGAN: GANの導入, より自然な高解像度画像を実現

- SRResNetの提案
  - 評価指標: PSNRはSRGANよりも良い結果.
  - 平均二乗誤差 (Mean Squared Error: MSE)をロスとしてトレーニング
    - 復元画像が自然画像の平均的なものになる → テクスチャや高周波成分の復元 ×
- SRGANの提案
  - SRResNetをGeneratorとして利用
  - Mean opinion score (MOS)テストにおいてSRResNetよりも良い結果.
    - MOSテスト: 26人のテスターによって行われたクオリティ評価



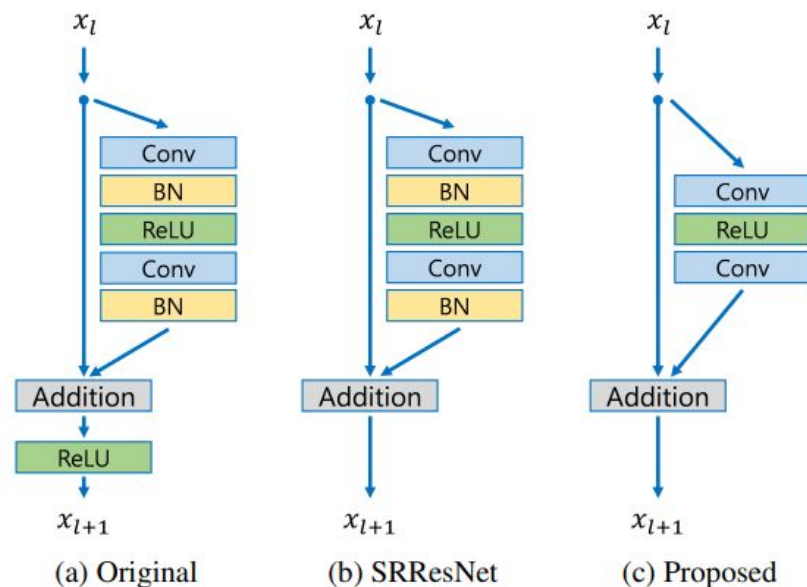
Set	SRResNet-		SRGAN-		
	MSE	VGG22	MSE	VGG22	VGG54
PSNR	32.05	30.51	30.64	29.84	29.40
SSIM	0.9019	0.8803	0.8701	0.8468	0.8472
MOS	3.37	3.46	3.77	3.78	3.58
<b>Set14</b>					
PSNR	28.49	27.19	26.92	26.44	26.02
SSIM	0.8184	0.7807	0.7611	0.7518	0.7397
MOS	2.98	3.15*	3.43	3.57	3.72*

会議 : CVPRW 2017

著者 : Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah and Kyoung Mu Lee

## EDSR/MDSR: SRResNetの改良・深層化, NTIRE 2017で優勝

- Enhanced Deep Super-Resolution Network (EDSR)
  - SRResNetからBatch Normalization (BN) 層を削除
    - BN層がなくても十分収束するため.
- Multi-scale Deep Super-Resolution system (MDSR)
  - 異なるスケールに対応したResBlockをアーキテクチャに組み込むことで, 単一モデルで異なるスケールのSRが可能
- EDSR, MDSRに対するself-ensembleの適用で性能向上
  - 入力画像の左右反転, 回転などの画像を用意
  - それぞれの復元画像を出力
  - 向きを揃えて平均を取る



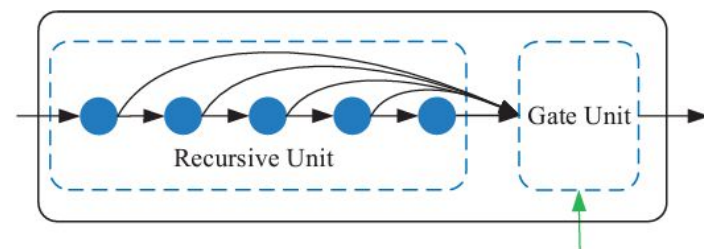


会議 : ICCV 2017

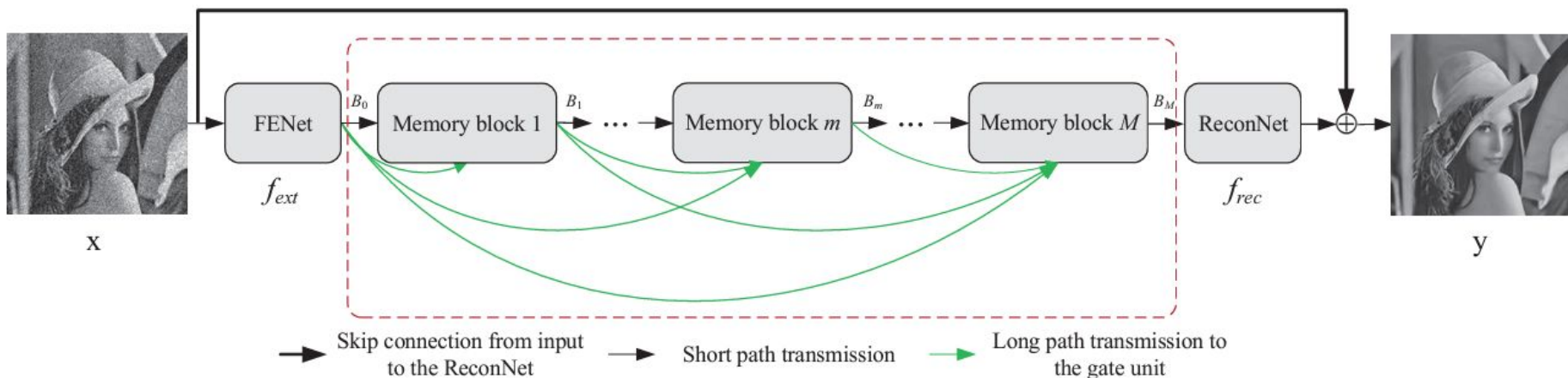
著者 : Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu

## MemNet: Gate Unitの追加により, 情報の長期記憶を実現

- 人の記憶の持続性に着目
- MemNet: memory blockを複数接続して構成
  - memory block: 残差接続とゲートユニット
- 深層のGate Unitまで情報を伝達.
- 集約した情報をReconstruction Net (ReconNet)に入力, 復元画像を出力.



(c) Proposed memory block



会議 : ICCV 2017

著者 : Tong Tong, Gen Li, Xiejie Liu and Qinquan Gao

## SRDenseNet: dense skip connectionsの採用, 深層CNNが高い再構成性能を達成することを示唆

- dense skip connectionにより, 低いレベルから高いレベルの特徴までを利用して画像復元において有効的に活用できる情報を集約できる.
- Dense block: 8つの畳み込み層で構成
  - growth rateが16(4倍の拡大率)なら, 出力は128次元の特徴マップとなる.

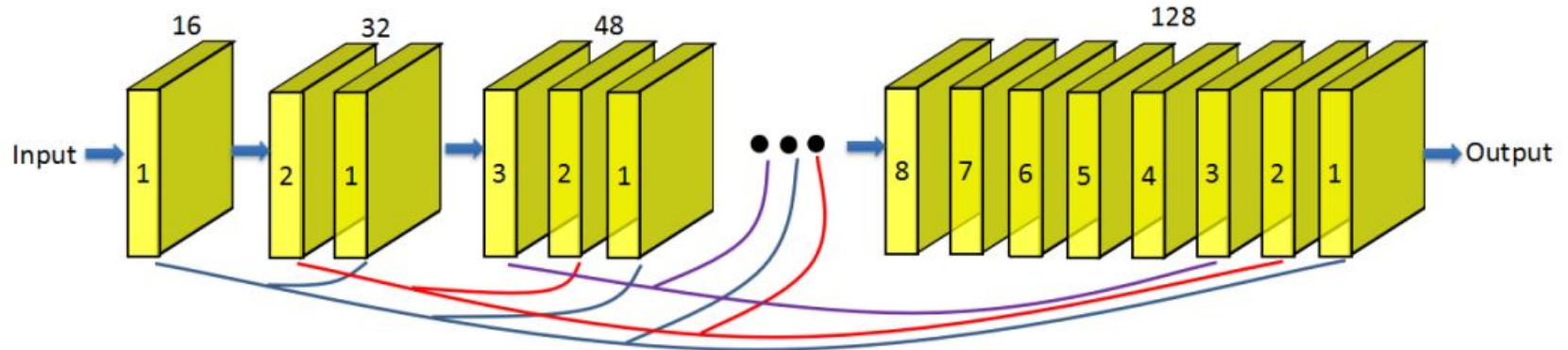


Figure 2. The structure of one DenseNet block. Each block consists of 8 convolution layers. The growth rate is set to 16 and the output of each block has 128 feature maps.

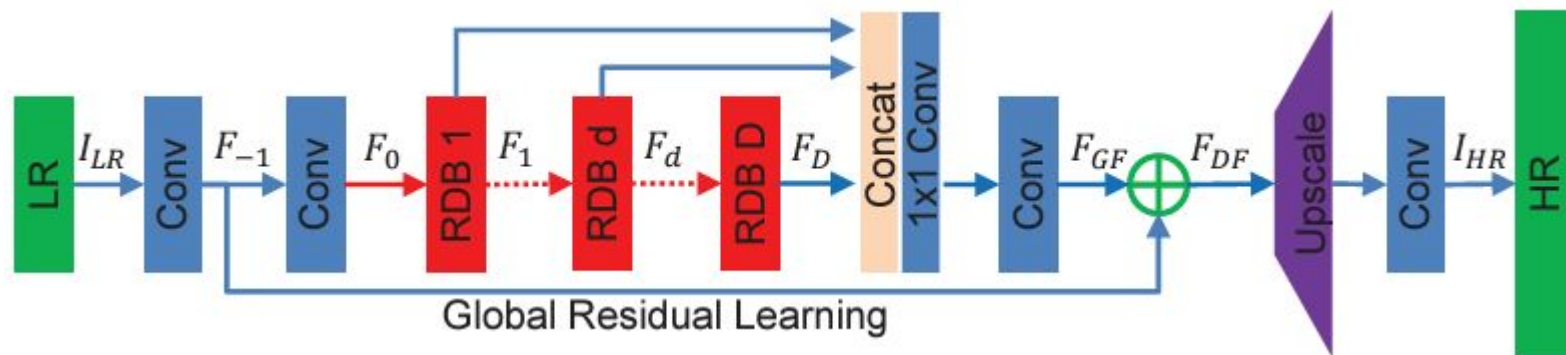
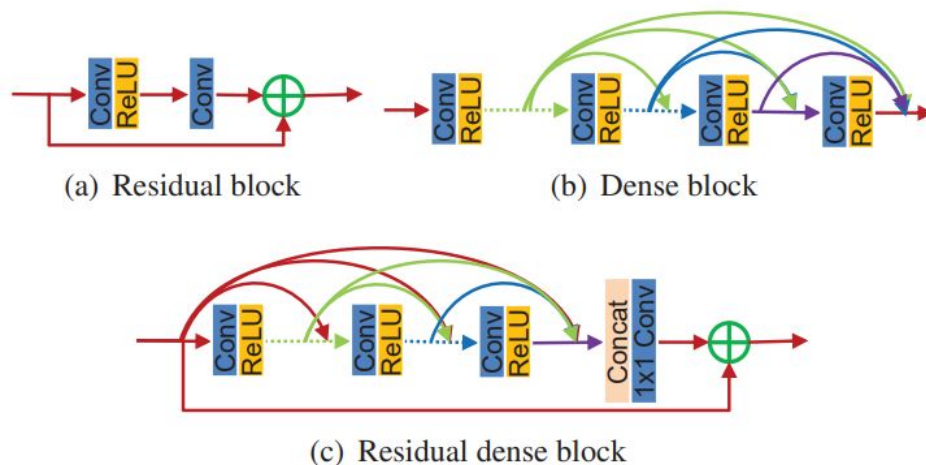
# Residual dense network for image super-resolution

会議 : CVPR 2018

著者 : Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong and Yun Fu

## RDN: Residual Dense Blockの追加, local情報の連続記憶を可能に.

- Residual Dense Block (RDB)
  - Dense blockにResidual blockを追加
    - block前後のサイズを合わせるため  
DenseNetの出力に $1 \times 1$  Convを追加
- 複数のRDBを含めたアーキテクチャを設計
  - local, global情報ともに有効的に  
集められることが強み

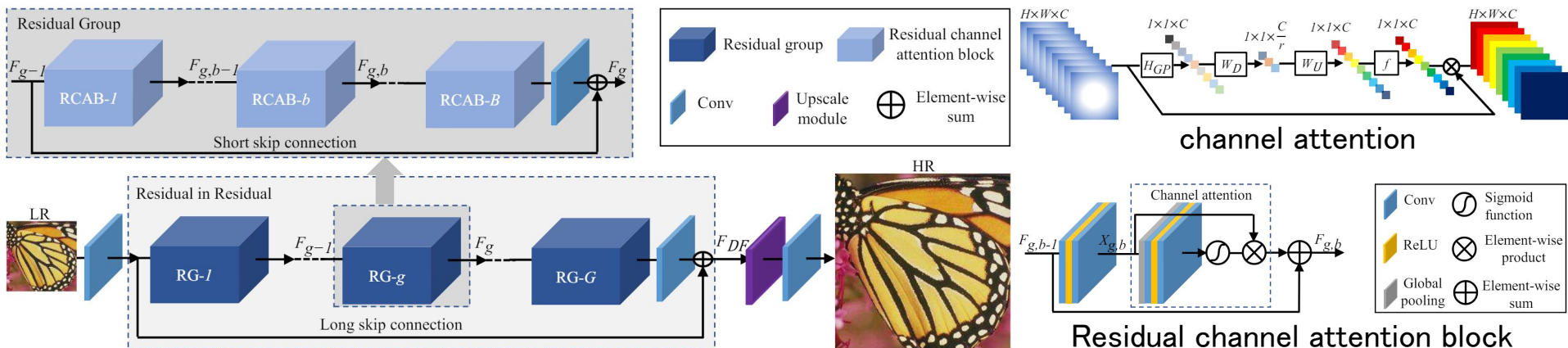


会議 : ECCV 2018

著者 : Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu

## RCAN: 残差ブロックとチャンネルアテンションにより深いネットワークを実現

- 400層以上の深い残差チャンネルアテンションネットワーク(RCAN)を提案
- RIR(Residual in Residual)により深いネットワークを訓練可能に
  - G個の残差グループと長いスキップ接続を含む(論文ではG=10)
  - スキップ接続を通じて低周波情報を保存
- チャンネルアテンションにより特徴チャンネル間の相互依存関係を考慮
- チャンネルアテンションを残差ブロックに統合したResidual Channel Attention Blockを提案
- SRタスクにおける深いネットワークとCAの有効性を定量的・視覚的に示した



Residual channel attention block

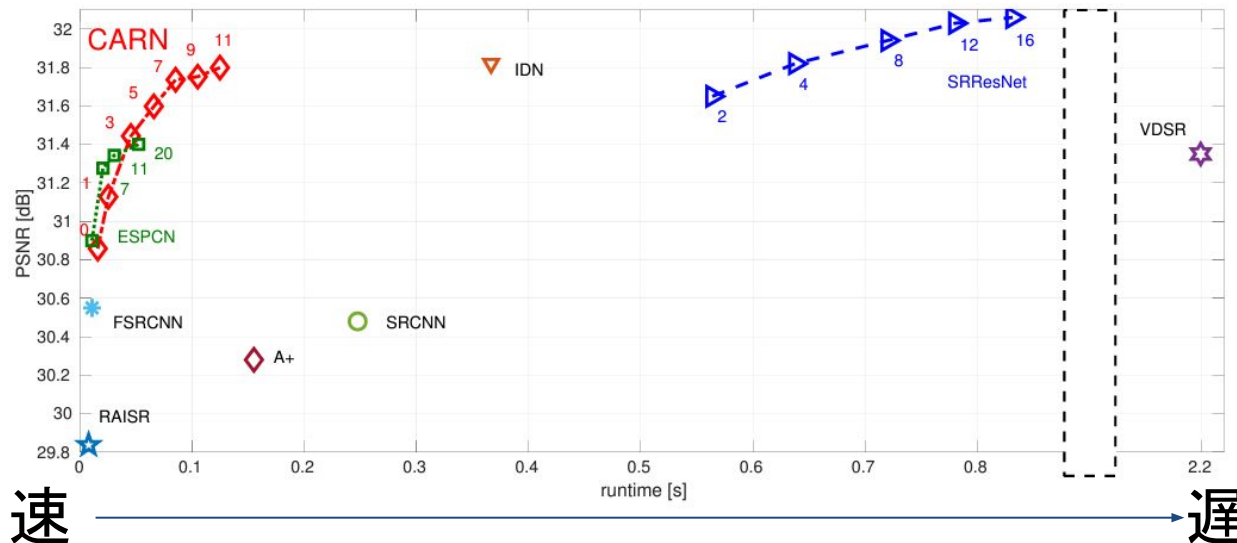
# CARN: Convolutional Anchored Regression Network for Fast and Accurate Single Image Super-Resolution

会議 : ECCVW 2018

著者 : Yawei Li, Eirikur Agustsson, Shuhang Gu, Radu Timofte, and Luc Van Gool

## CARN: 抽出した特徴をアンカーで分類後に回帰, 計算高速化

- 局所回帰の手法(locally linear regression methods)を参考としてregression blocks追加
- A+やARNといった局所回帰の手法とは異なり, ハンドクラフトの特徴抽出器をCNNで代用
- End2endのCNNモデルによるSRの計算時間と復元クオリティのトレードオフを大幅に改善



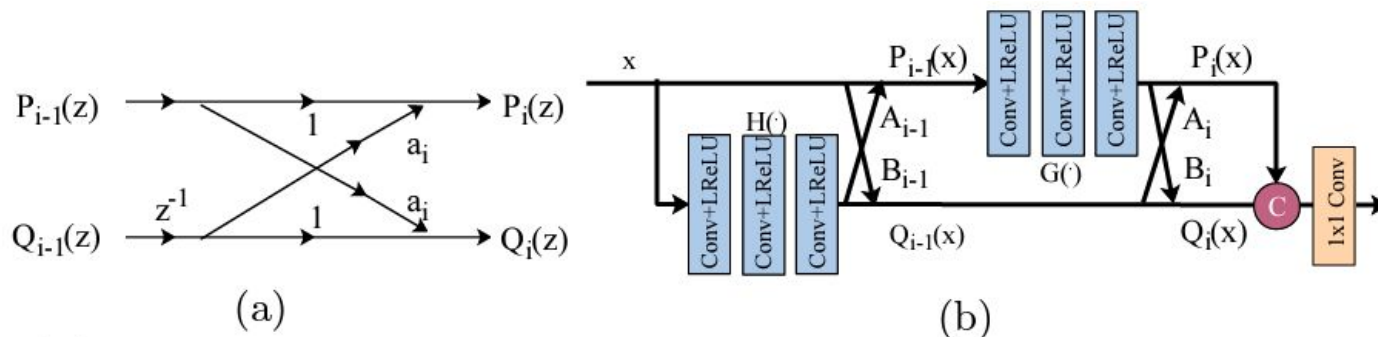
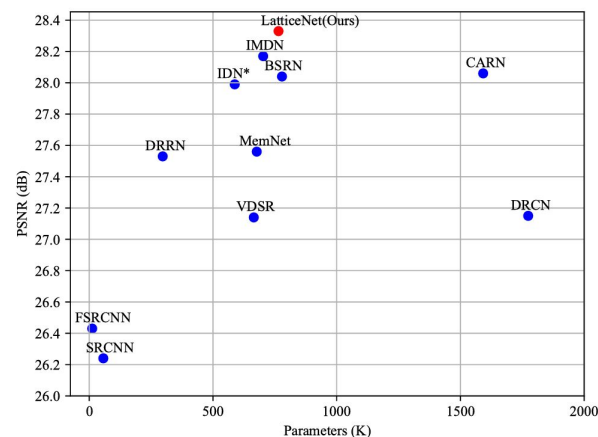


会議 : ECCV 2020

著者 : Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu

## LatticeNet: ラティスブロックを使用して残差ブロックの出力を効率的に結合

- 少量パラメータで追加できるLattice blockによって既存のSRモデルの性能を向上できることを示した
- Lattice blockを基準として設計したLatticeNetは少量パラメータで既存SRモデルを凌駕する性能を発揮



**Fig. 2.** (a) The basic unit of a standard lattice structure for a 2-channel filter bank. (b) The structure of proposed lattice block

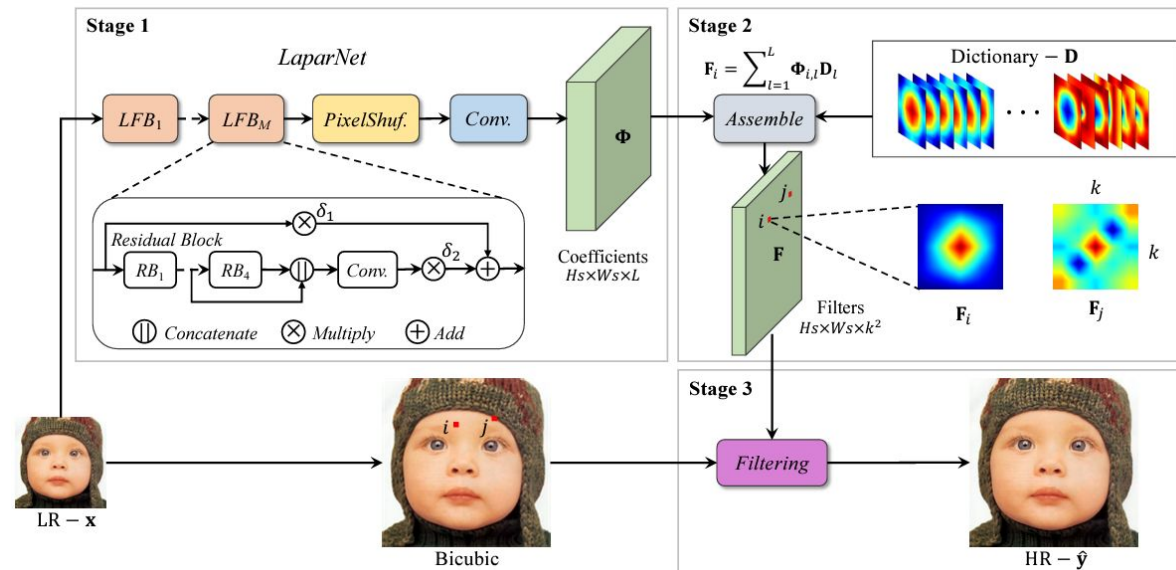
# LAPAR: Linearly-Assembled Pixel-Adaptive Regression Network for Single Image Super-Resolution and Beyond

会議 : NeurIPS 2020

著者 : Wenbo Li1, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, Jiaya Jia

## LAPAR: 事前定義された複数のフィルタ基底の辞書を使用

- Stage 1: 軽いResidual networkを用いて係数(Coefficients)を算出
- Stage 2: Stage 1の係数と辞書を用いてフィルタFの行列を算出
- Stage 3: Bicubic upsampling画像に対してStage 2のフィルタを適用してHR画像実現
- LAPARのアイデアはimage denoisingやJPEG image deblockingなどの、超解像以外の画像復元タスクにも適用可能





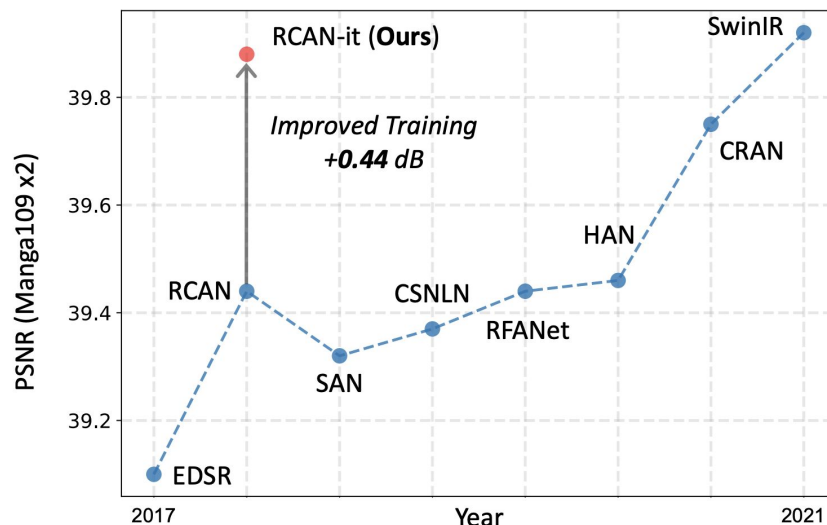
会議 : arXiv 2022

著者 : Zudi Lin, Prateek Garg, Atmadeep Banerjee, Salma Abdel Magid, Deqing Sun, Yulun Zhang, Luc Van Gool, Donglai Wei, Hanspeter Pfister

## RCAN-it: 学習戦略と最小限のアーキテクチャ改善により, RCANの性能を向上

- モデルの進化に伴い学習設定も改善されているが, 性能比較時に古いモデルの学習設定は更新されないため, 不公平な比較が生じていることに注目した論文
- 400以上の層で構成されておりポテンシャルが高いRCANの学習設定をアップグレードして性能比較
  - DF2Kの学習時にunderfittingしていることに注目し, 学習時のiteration数を増加
  - バッチ数, optimizer, 学習率スケジューリングの改善および分散学習の導入より 学習時間を短縮
  - 活性化関数をReLUからSiLUに変更

Method		Set5	Set14	B100	Urban100	Manga109
RCAN [50]	x2	38.27	34.12	32.41	33.34	39.44
RCAN-it	x2	<b>38.37</b>	<b>34.49</b>	<b>32.48</b>	<b>33.62</b>	<b>39.88</b>
RCAN [50]	x4	32.63	28.87	27.77	26.82	31.22
RCAN-it	x4	<b>32.69</b>	<b>28.99</b>	<b>27.87</b>	<b>27.16</b>	<b>31.78</b>

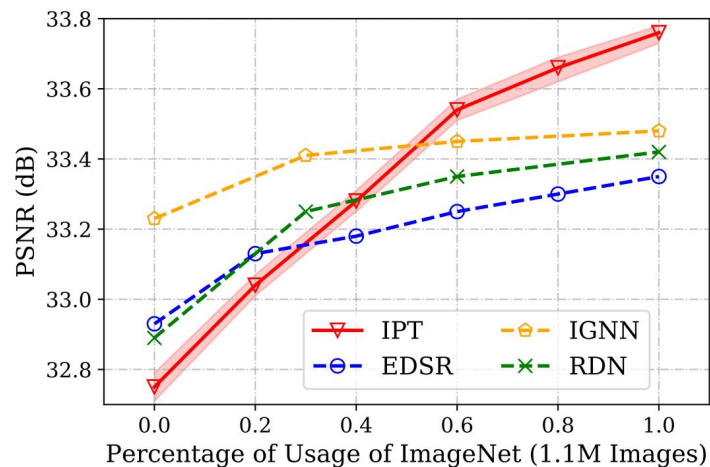
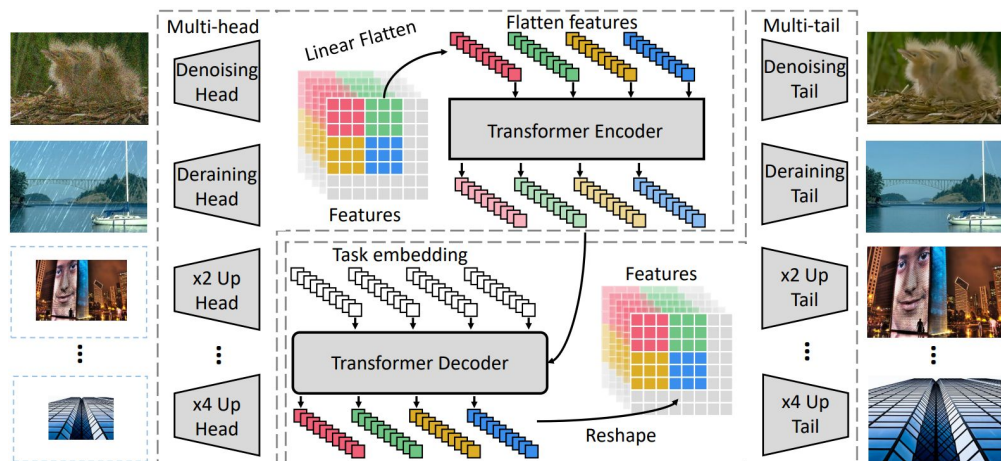


会議 : CVPR 2021

著者 : Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao

## IPT: SR, Denoising, Deraining に適用可能なTransformer系モデル

- 事前学習にImageNet1kを利用
  - 多様なdegradationに対応するために対照学習を行う
  - 同一画像からのpatchの埋め込みは近づけ, 異なる画像に対しては遠ざける
  - Transformer系のモデルは大規模データでの事前学習によりCNN系のモデルの精度を上回る
- 複数のheadとtailで複数タスクに対応
  - fine-tuningにより新たなタスクにも適用可能

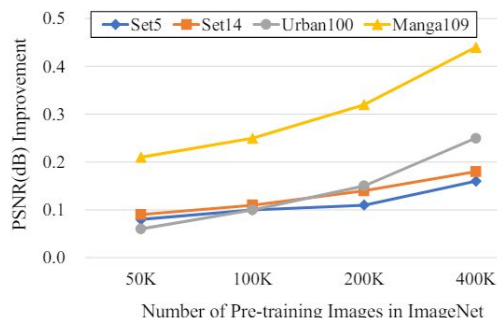
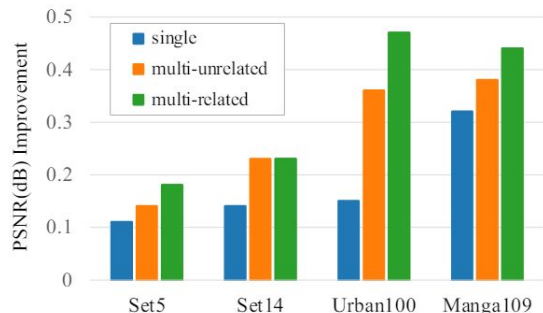


会議 : IJCAI 2023

著者 : Wenbo Li, Xin Lu, Shengju Qian, Jiangbo Lu

## EDT : low-level visionタスクに対する事前学習の効果を検証

- low-levelな画像処理タスクに対する事前学習の効果を調査
  - 超解像では, より局所的な情報を高次の層に導入することができるようになる
  - ノイズ除去では内部の特徴にほとんど影響を与えない
- 単一タスク, マルチタスク(SRとノイズ除去), マルチ関連タスクを事前学習として実験
  - マルチ関連タスク(異なる複数のスケールのSR)事前学習が効率的であると示した
  - マルチ関連タスクではデータセットの枚数が少なくても効果があった
- 効率的に調査するためにEDTアーキテクチャを提案
- 



Model	Data	Set5	Set14	Urban100	Manga109
EDT-B	0	38.45	34.57	33.80	39.93
EDT-B <sup>†</sup>	50K	38.53	34.66	33.86	40.14
EDT-B <sup>†</sup>	100K	38.55	34.68	33.90	40.18
EDT-B <sup>†</sup>	200K	38.56	34.71	33.95	40.25
EDT-B <sup>†</sup>	400K	38.61	34.75	34.05	<b>40.37</b>
EDT-B*	200K	<b>38.63</b>	<b>34.80</b>	<b>34.27</b>	<b>40.37</b>

\* はマルチ関連タスクの事前学習を行ったことを示す

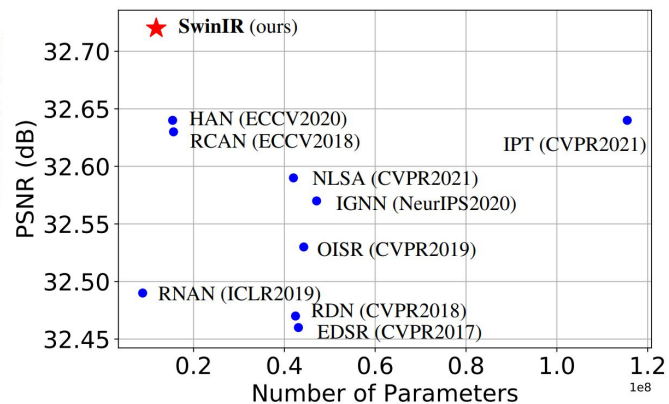
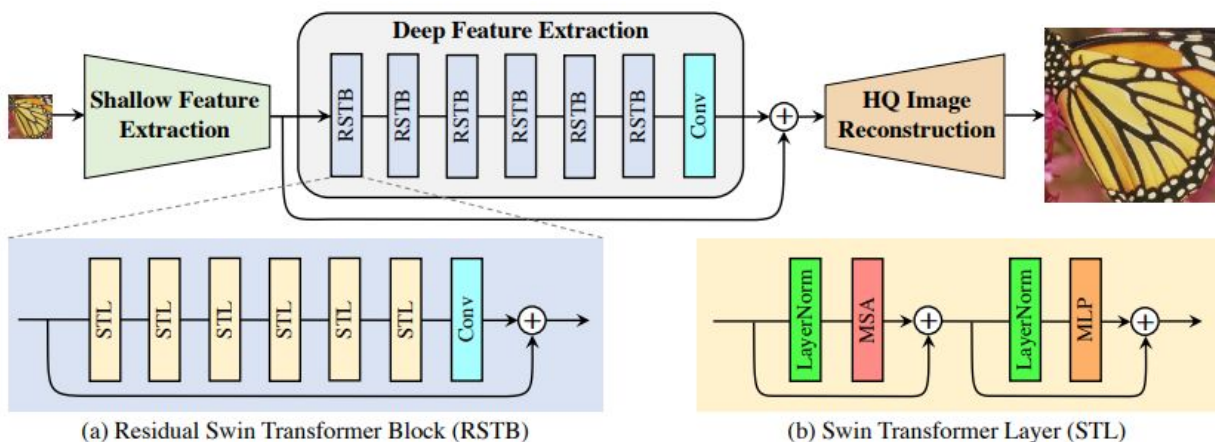
# SwinIR: Image Restoration Using Swin Transformer

会議 : ICCV Workshop 2021

著者 : Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte

## SwinIR: Swin TransformerをSRに導入

- Shallow/Deep Feature Extraction, HQ Image Reconstructionの3段構成
  - Deep Feature ExtractionにSwin Transformerの機構を導入
  - residual connectionにより異なるレベルの特徴を集約
- IPTと比べて省パラメータ, 高パフォーマンスを達成
  - IPTではCNNベースのモデルを上回るのに多くのデータを要していたのに対し, SwinIRでは同量のデータでCNN系のモデルの性能を上回った



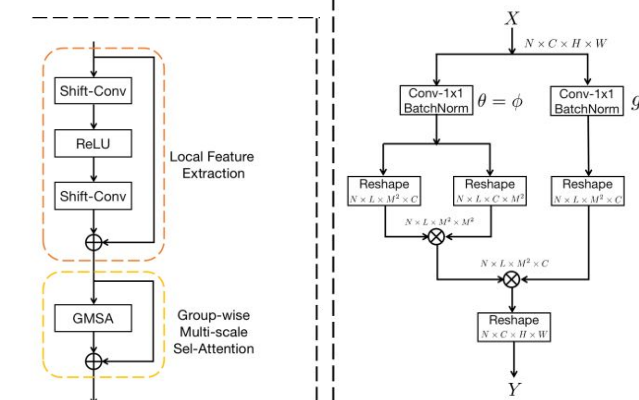
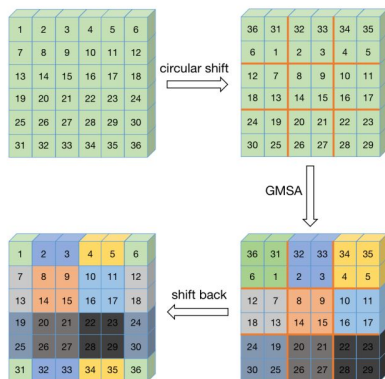
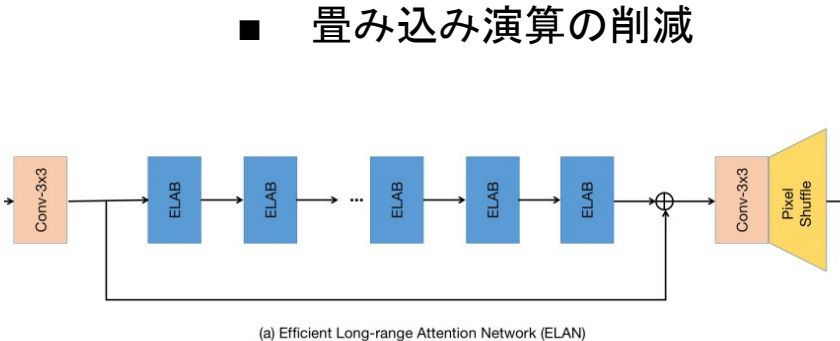
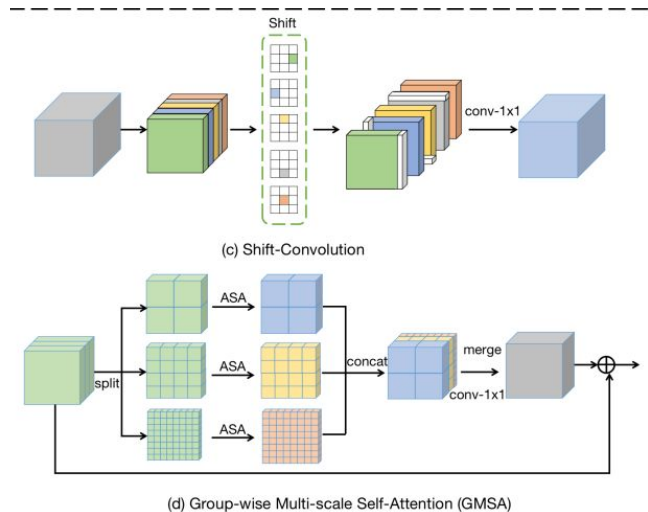
<https://github.com/JingyunLiang/SwinIR>

会議 : ECCV 2022

著者 : Xindong Zhang, Hui Zeng, Shi Guo, Lei Zhang

## ELAN: シフト畳み込みとGMSAモジュールによるELABを提案

- 既存のモデルの構造を再考し, 計算コストを抑えたシンプルな機構を提案
- ELAB :
  - shift-conv : 計算コストを抑えて局所的な構造を抽出
  - GMSA : 入力特徴を複数グループに分割し, それぞれのグループについてSelf Attentionをとることで長距離の依存性を獲得
  - ASA : 推論時のSelf Attentionの速度を改善
    - Batch Normの採用
    - 畳み込み演算の削減



(c) Shifted Window for Consecutive ELABs

(b) Efficient Long-range Attention Block (ELAB)

(e) Accelerated Self-Attention (ASA)

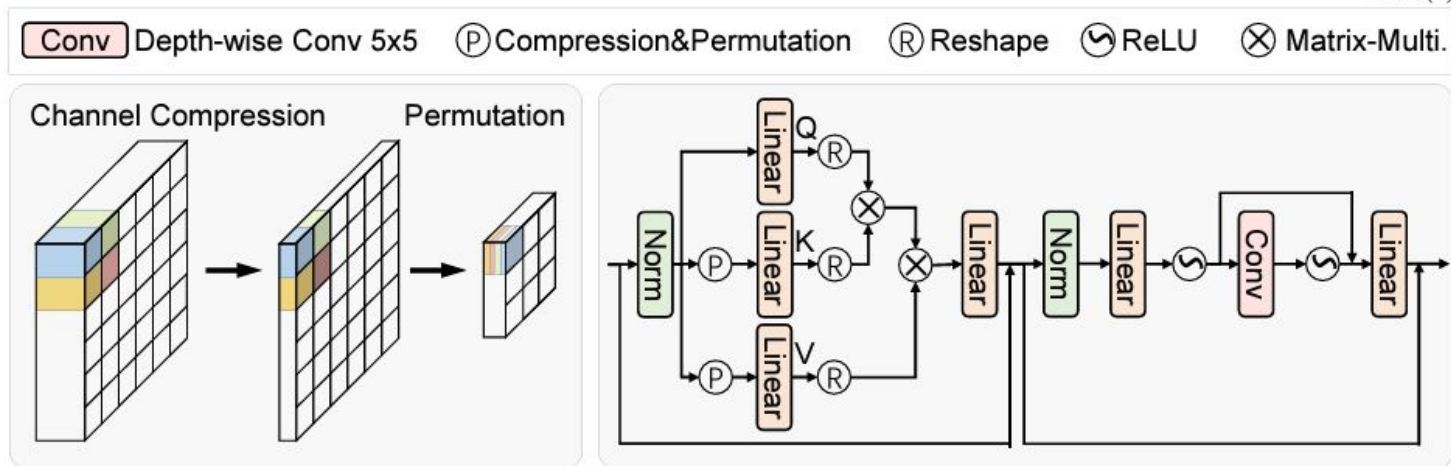
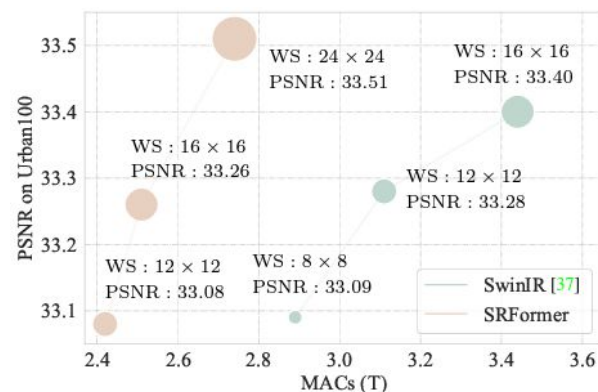


会議 : ICCV 2023

著者 : Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, Qibin Hou

## SRFormer: チャンネル方向, 空間方向の情報をバランスよく扱える permuted self-attention 機構を提供

- permuted self-attention (PSA)は既存の window self-attentionベースのSRモデルに簡単に適用可能
- SwinIRよりも少量パラメータ, 少量計算コストで, Urban100において0.46 dB高いスコアを達成

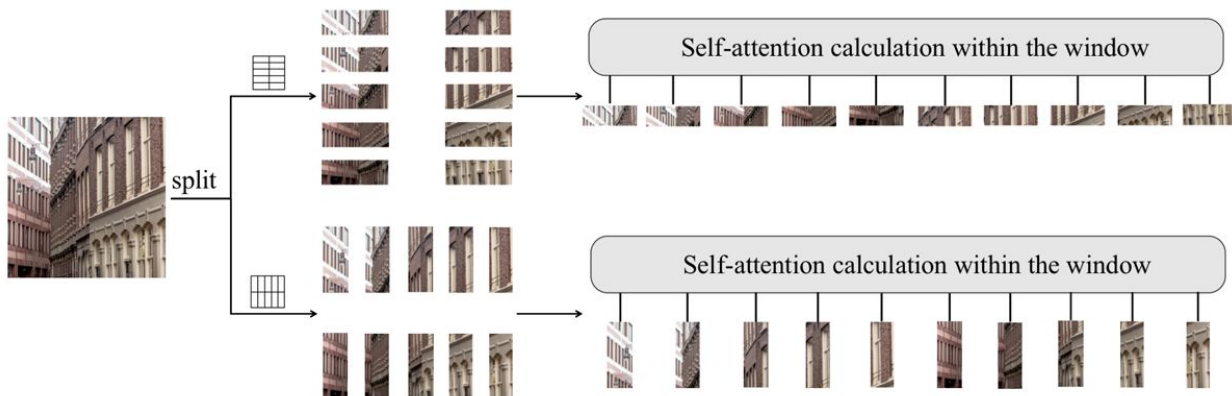
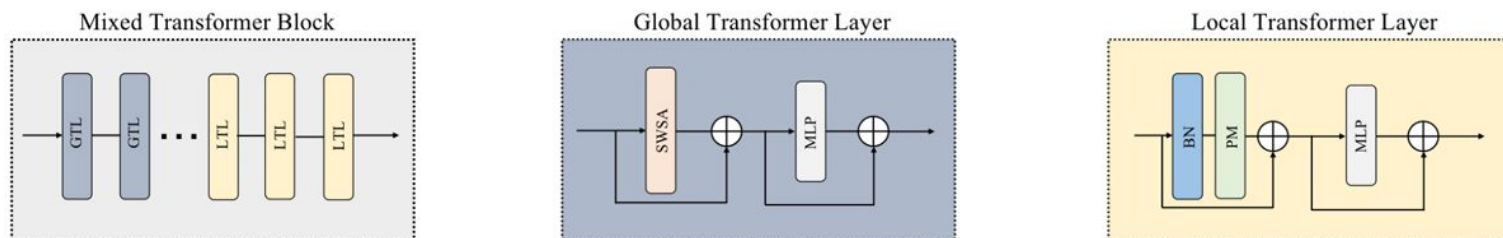


会議 : Engineering Applications of Artificial Intelligence 2024

著者 : Ling Zheng, Jincheng Zhu, Jinpeng Shi, Shizhuang Weng

## EMT: 局所的かつ複雑な画像要素の復元の難しさを克服したTransfer

- Efficient Mixed Transformer (EMT)の提案
  - Mixed Transformer Block (MTB)の導入
    - パラメータは増加しない
    - 計算時間 (floating-point operations) も増加しない



- striped window for SA (SWSA)
  - 局所／大域的な情報を扱う性能向上
  - window self-attentionベースのSRモデルに適用可能

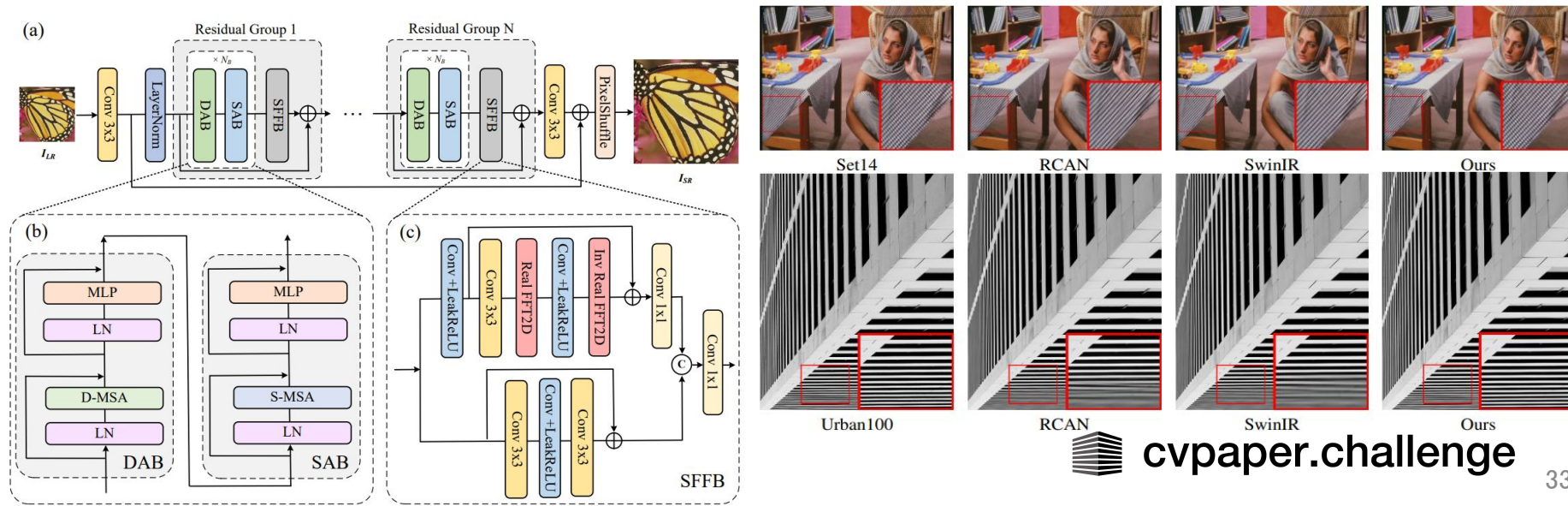


会議 : CVPR 2023

著者 : Qiang Zhu, Pengfei Li, Qianhui Li

## ARFFT: 受容野を画像全体に拡張

- spatial-frequency fusion block(SFFB) を提案
  - 周波数領域と空間領域の特徴を利用
  - 受容野を画像全体に拡大しモデルの表現能力を拡張
- Dense attention block(DAB)とSparse AB(SAB)により局所と非局所の情報抽出
  - D-MSAで近傍のトークンと, S-MSAで離れた位置のトークンとの相互関係を学習
- パッチサイズを徐々に変更して学習を行う手法を提案
- FFTLossを使用

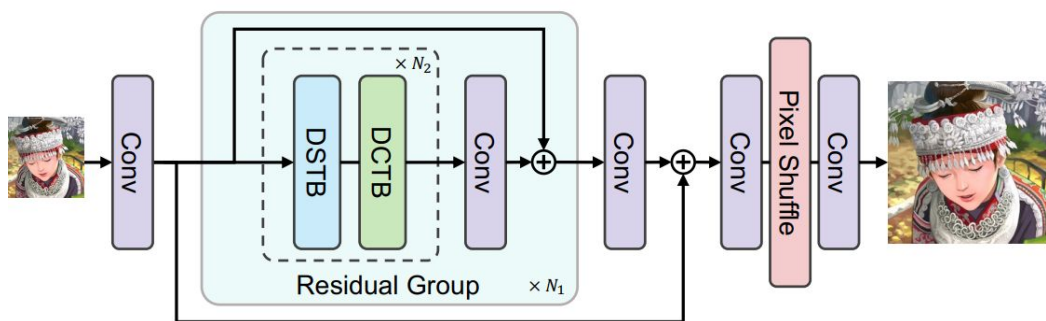


会議 : ICCV 2023

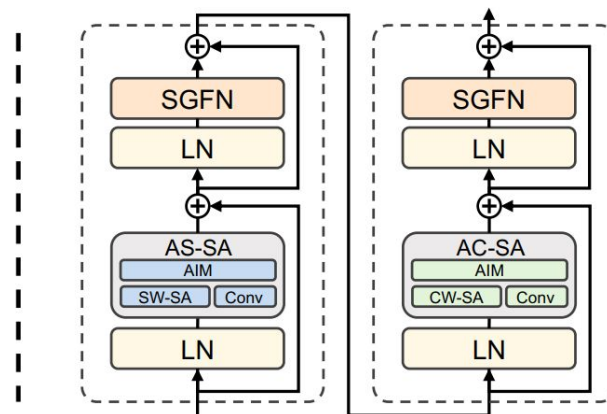
著者 : Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, Fisher Yu

## DAT:空間とチャネルの次元を横断してAttentionを集約

- Dual Aggregation Transformer Block
  - 空間方向(DSTB)とチャネル方向(DCTB)のTransformerブロックを交互に適用することでブロック間とブロック内で空間特徴とチャネル特徴を集約
- adaptive interaction module (AIM)
  - self-attentionと畳み込み層の特徴を効果的に集約
- spatial-gate feed-forward network (SGFN)
  - FFNに非線形空間情報を導入



(a) DAT



(b) DSTB

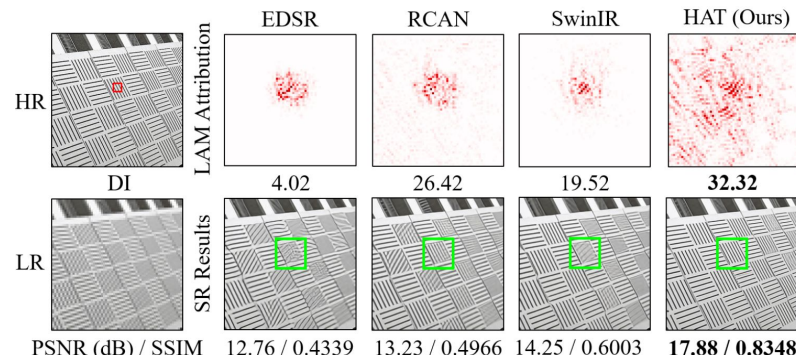
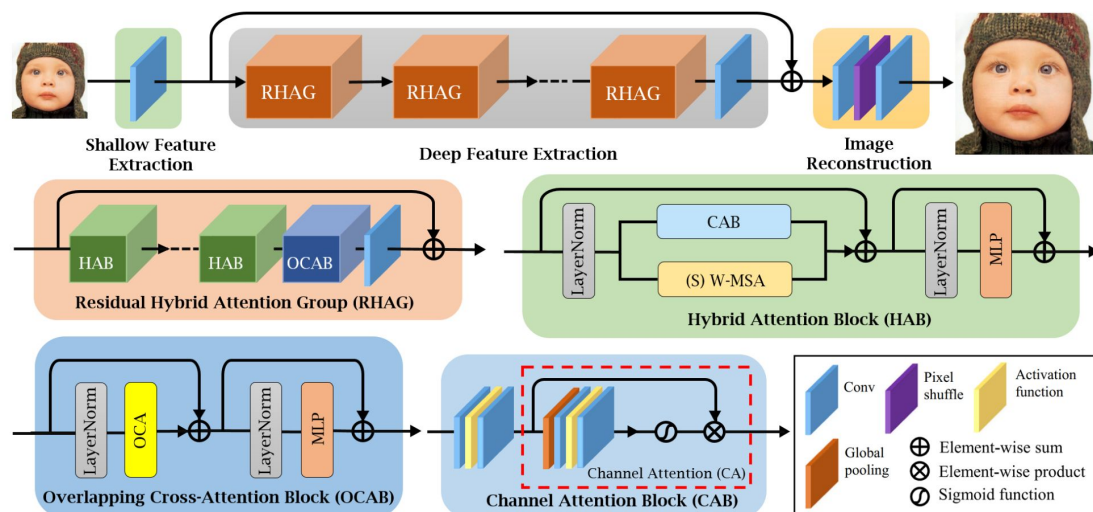
(c) DCTB

会議 : CVPR 2023

著者 : Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, Chao Dong

## HAT: より多くの入力ピクセルを再構成に利用するモデルを提案

- チャンネルアテンションとウィンドウベースのセルフアテンションを並列に導入
  - チャンネル注意ブロック(CAB)によりグローバル情報を抽出し, 多くのピクセルを活性化
  - ウィンドウ内で計算される自己注意(W-MSA)で局所情報を抽出
- オーバーラッピングクロスアテンションブロック(OCAB)を導入してウィンドウ間を接続
  - 隣接ウィンドウ特徴間の相互作用を強化
- 同タスクでImageNetを用いた事前学習を行うことによりモデルのポテンシャルを引き出す
- 当時のSoTaを1dB以上上回る性能を達成



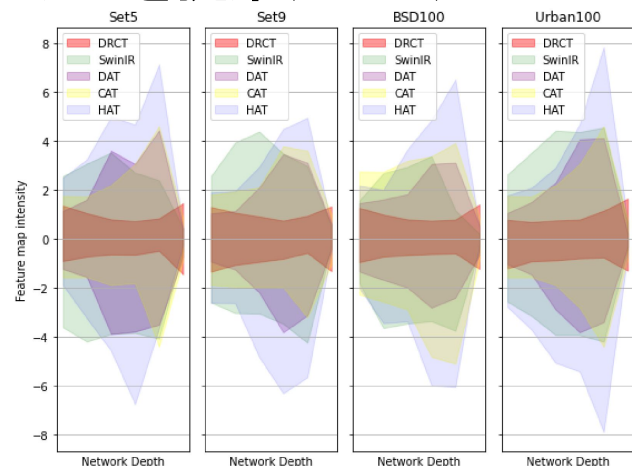
再構成に利用されている入力ピクセルの可視化

会議 : CVPR Workshop 2024

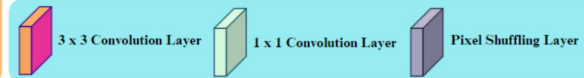
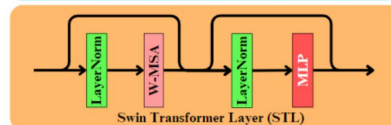
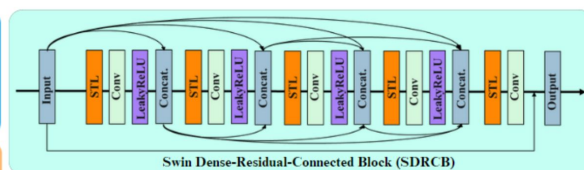
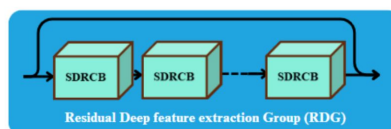
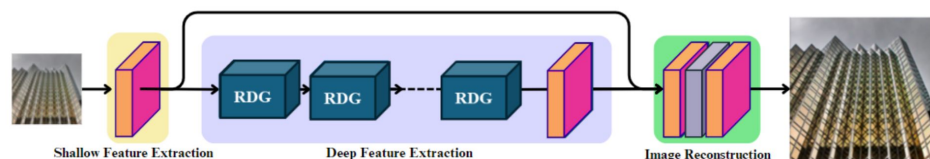
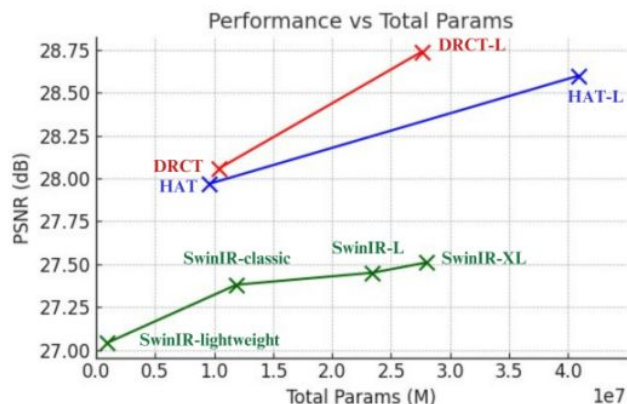
著者 : Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou

## DRCT: 多層ネットワークによる情報損失を抑制したモデルを提案(SOTA)

- 既存のベンチマークにおいてx2, x3, x4SOTA
- 従来のモデルではネットワークの末尾で特徴マップの急激な変化が生じており, 空間情報の損失を伴う可能性有があった
  - Residual Deep feature extraction Group(RDG)
    - SDRCBを複数重ねることで, 異なるスケールの情報を統合し, 残差接続により情報の損失を防止/前方伝播プロセスの安定化
  - Swin-Dense-Residual-Connected Block(SDRCB)
    - シフトウィンドウ自己注意を使用して適応的な受容野を獲得
- Progressive Training Strategy(PTS)をトレーニング戦略として採用



networkの末尾で特徴マップの強度が低下





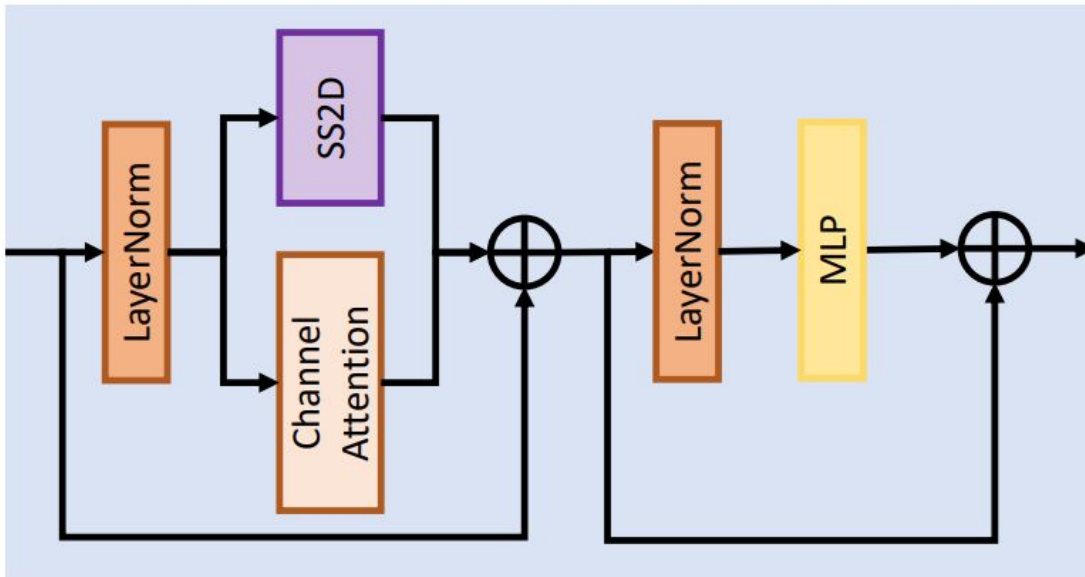
会議 : CVPRW 2024

著者 : Hongyuan Yu, Cheng Wan, Yuxin Hong, Zhijuan Huang, Yajun Zou, Yuan Huang, Jiamin Lin, Bingnan Han, Xianyu Guan, Yongsheng Yu, Daoan Zhang, Xuanwu Yin, Kunlong Zuo

## MambaSR: Mamba とTransformer を組み合わせた SSFormer blockを提案

- NTIRE 2024 Challenge on Image Super-Resolution (×4)にて最高性能を達成したモデル
  - 状態空間モデルを用いたTransformer(State Space Transformer: SSFormer)の導入
- HATをベースとして, Hybrid Attention BlocksをSSFormerによって置換

Team	Rank	PSNR (primary)	SSIM
XiaomiMM	1	31.94	0.8778
SUPSR	2	31.41	0.8711
UCAS-SCST	3	31.28	0.8666
SYSU-SR	4	31.19	0.8660
Jasmine	5	31.18	0.8665
ACVLAB	6	31.18	0.8655
mandalinadagi	7	31.13	0.8648
SKDADDYS	8	31.11	0.8643
KLETech-CEVI	9	31.03	0.8633
SVNIT-NTNU	10	31.03	0.8633
ResoRevolution	11	31.01	0.8647
BetterSR	12	30.97	0.8621
Contrast	13	30.69	0.8563
BFU-SR	14	30.55	0.8560
SCU-VIP-LAB	15	29.78	0.8506
Nudter	16	30.17	0.8446
JNU-620	17	30.43	0.8426
LVGroup-HFUT	18	29.98	0.8380
Uniud	19	29.97	0.8440
SVNIT-NTNU-1	20	29.34	0.8199





会議 : arXiv 2024

著者 : Xiaoyan Lei, Wenlong Zhang, Weifeng Cao

## DVMSR: ViMと知識蒸留により高速高品質のSRを実現

- Vision Mamba Moduleと残差接続を組み込んだブロックを提案
  - Mambaの長距離モデリングの能力と効率的かつ複雑な計算性能をSRタスクに利用する試み  
知識の蒸留によりモデルを軽量化
- メタ話
  - Mambaの特徴である長距離の入カシーケンスに対して優れているという点と, 超解像の対象である高解像度画像の親和性が高いため, 超解像分野でのMambaベースのモデルがさらに発展する予感.

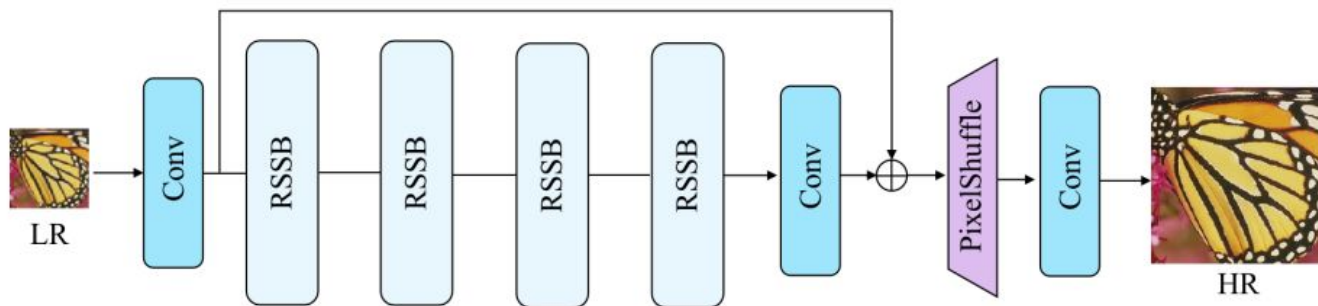


Figure 3. The overall network architecture of our DVMSR.

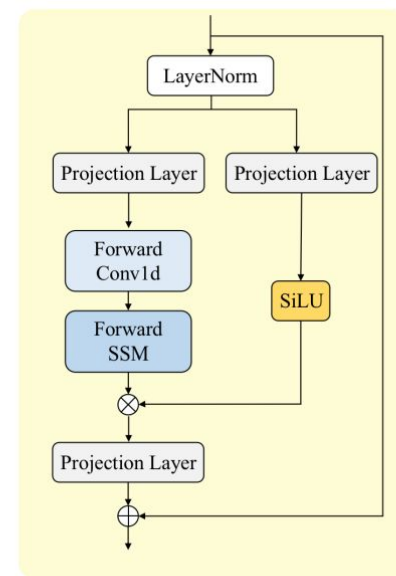
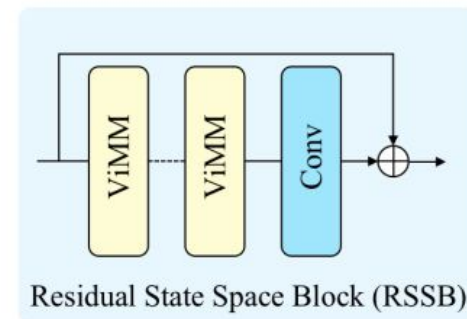


Figure 4. The structure of Vision Mamba Module(ViM).

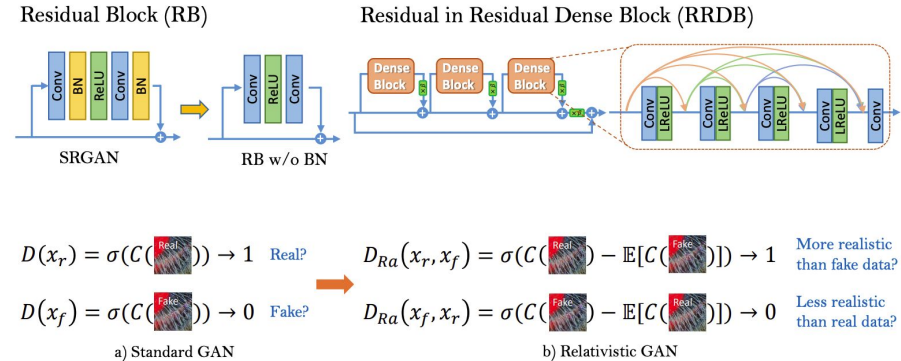
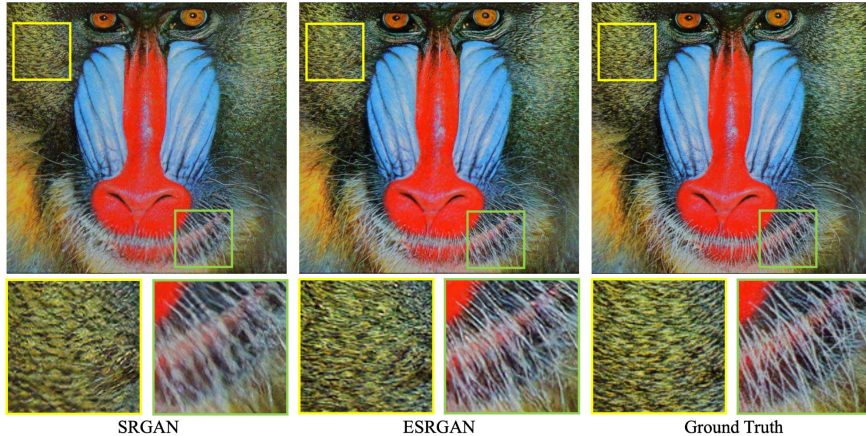


会議 : ECCV 2018 Workshop

著者 : Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy

## ESRGAN: SRGANの各要素をアップデートして精度向上

- SRGANのアーキテクチャ, adversarial loss, perceptual lossを見直すことで精度向上
  - Residual Block (RB)からResidual-in-Residual Dense Block (RRDB)へ変更
  - Discriminatorを実データと偽データの相対的な関係を考慮するRelativistic GANに変更
    - 相対的な関係にすることで学習が安定
    - エッジやテクスチャといった細部の表現を考慮した学習が可能
  - Perceptual lossをactivation後ではなく, activation前の特徴マップで算出
    - activation後はテクスチャや明度などの情報量が減少してしまうため

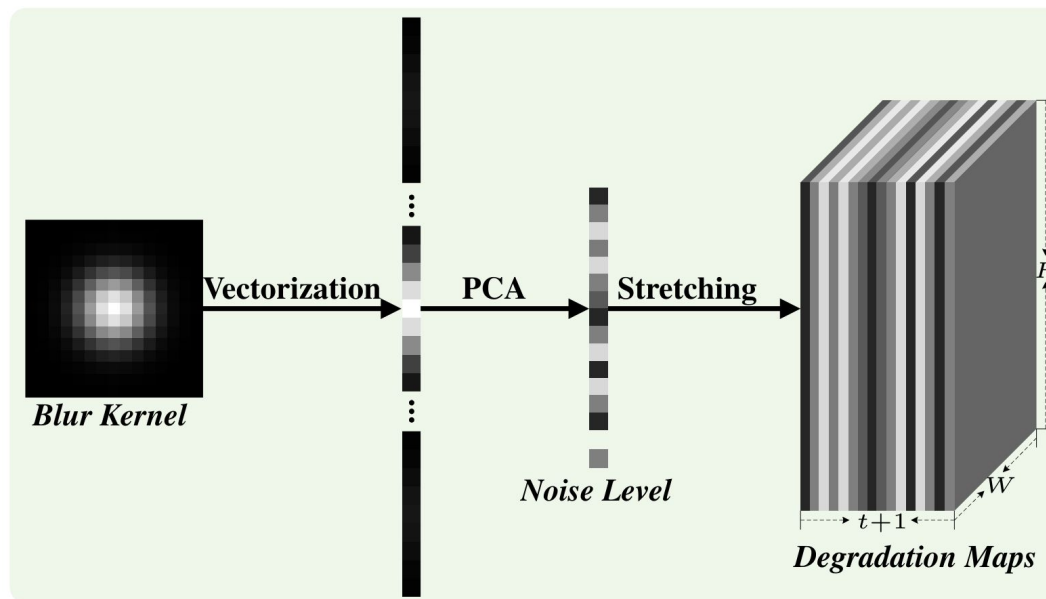
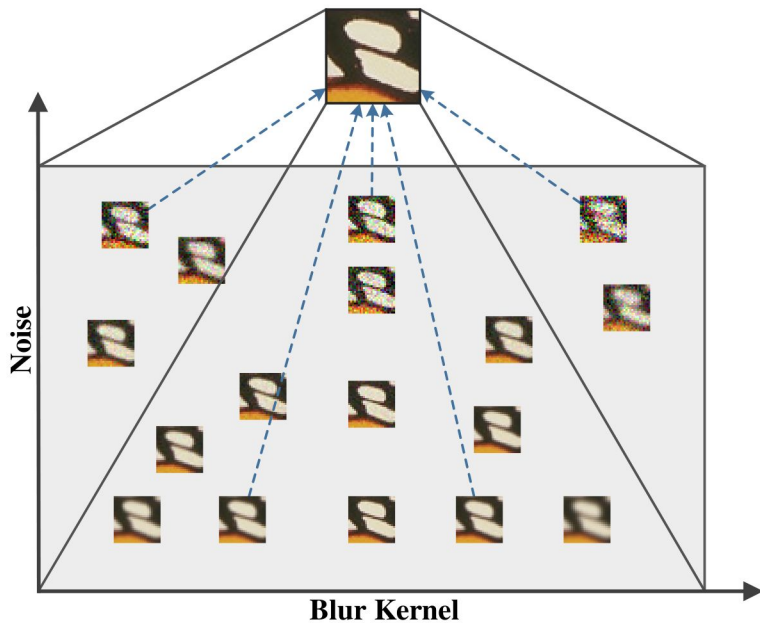


会議 : ECCV 2018 Workshop

著者 : Kai Zhang, Wangmeng Zuo, Lei Zhang

## 実際のLR画像に対処するため複数の劣化を組み合わせることを提案

- 次元の異なるblurとノイズを同時に扱うための次元ストレッチングを提案
  - Blurカーネルをベクトル化し, PCAを用いて $t$ 次元に圧縮
  - ノイズレベルを結合して $t+1$ 次元とし, 劣化マップを作成
- bicubic劣化において最先端の結果を達成
- 他の劣化や空間的に変化する劣化に対しても有利な性能を示した



会議 : ICCV 2017

著者 : Mehdi S. M. Sajjadi, Bernhard Scholkopf, Michael Hirsch

## 敵対的学習 + perceptual loss + texture loss で認知的精度を向上

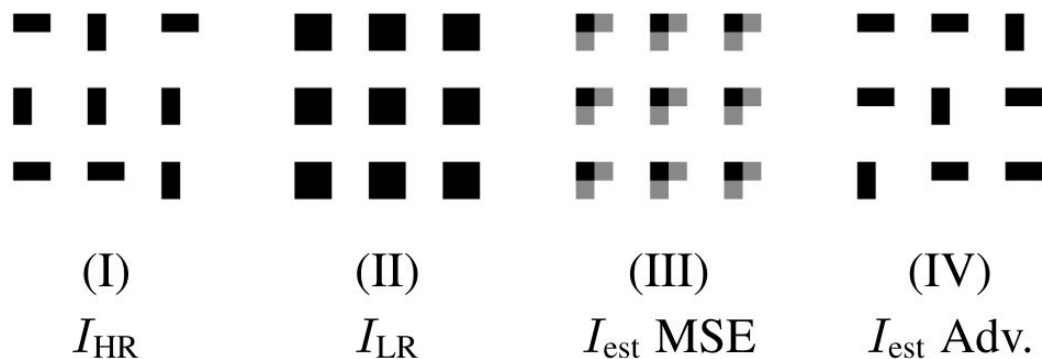
- pixel空間でMSEの最適化することによって画像の平滑化が生じる
  - pixel-wiseでは縮小時に失った高周波情報を平均的な値で埋めるため
- 畳み込み層のみを使用したFCNをベースとしたモデル
- 畳み込み層をアップサンプリング後に挿入することでアーチファクトを抑制
  - 畳み込み層によるアップサンプリングは格子状のアーチファクトが生じる
- $16 \times 16$ のpatch単位のstyle transfer lossにより局所的なテクスチャを学習



State of the art by PSNR



Our result

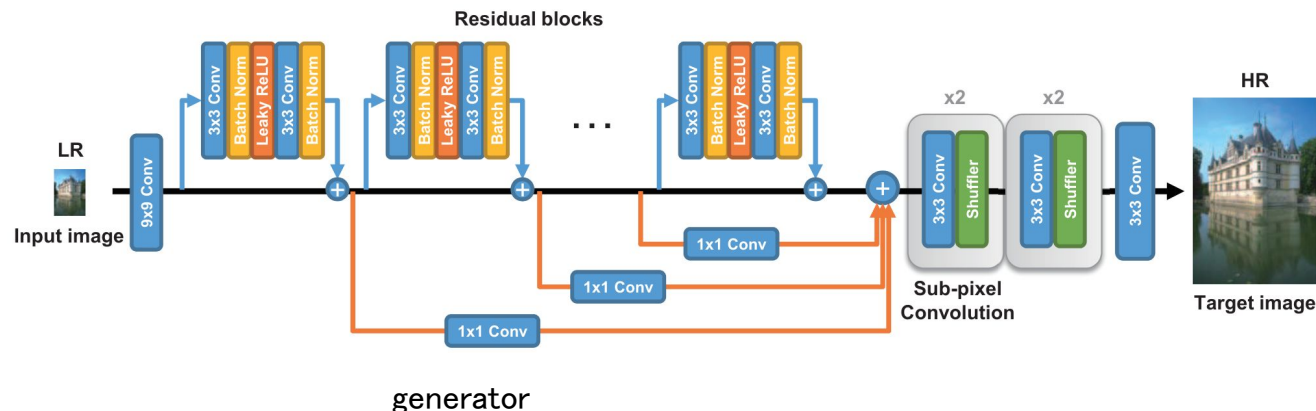


会議 : ECCV 2018

著者 : Seong-Jin Park, Hyeongseok Son, Sunghyun Cho, Ki-Sang Hong

## SRFeat : 特徴空間のDiscriminatorとgeneratorの改善により精度向上

- pixel空間だけでなく特徴空間へのdiscriminatorを追加
  - GANモデルによる復元画像にはpixel空間のdiscriminatorを騙すために無関係な高周波ノイズが含まれる
  - 特徴空間へのdiscriminatorの導入により, 認知的にもGTに近い画像を生成可能に
- long range skip connectionを用いたgenerator
  - 離れた層からの情報を効率的に伝搬
  - PSNRとSSIMを改善



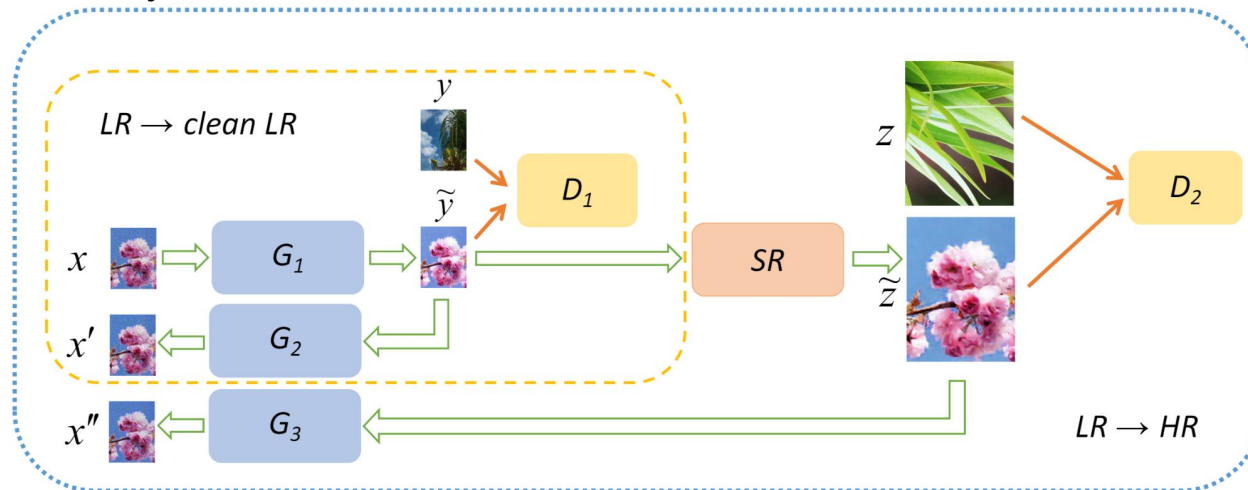


会議 : CVPRW 2018

著者 : Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, Liang Lin

## CinCGAN : ペアデータなしでノイズの多いLRからHRを生成

- 二つのCycleGANを用いたCycle-in-Cycle構造を提案
  - 二つ目のGANが一つ目のGANを覆っている構造
  - CycleGANでは入出力スケールを変更できないため、学習済みSRモデルを導入
- 1. 一つ目のCycleGANによりLRからnoise/blurを除去 ( $G_1, G_2, D_1$ )
- 2. 学習済み深層学習モデルでアップサンプリング (SR)
- 3. 二つ目のCycleGAN全体をfine-tune ( $G_1, SR, G_3, D_2$ )



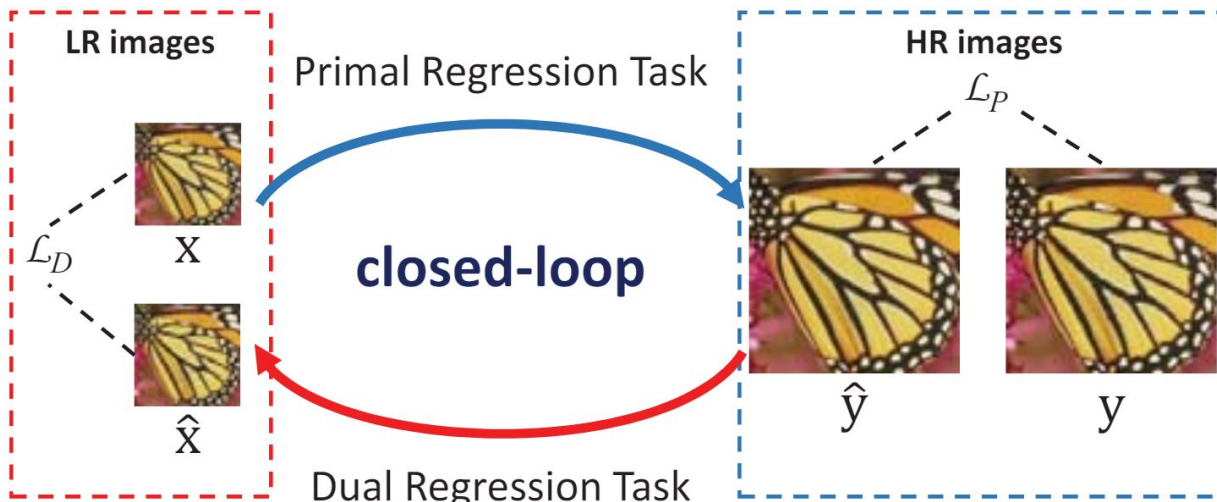
CinCGAN

会議 : CVPR 2020

著者 : Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhong Cao, Zeshuai Deng, Yanwu Xu, Mingkui Tan

## DRN : LR→HRの変換とHR→LRの変換を同時に学習

- ill-posed problemに対処するために画像の空間に制限を加える
  - LR→HRのマッピングが閉ループを形成できるようにする
  - LR→HRの変換とHR→LR変換を同時に学習
  - HR→LRの変換を学習することにより, 実世界の劣化の学習を行うことが出来る
- HRが与えられていない実世界の場合にも有効



<https://github.com/guoyongcs/DRN>

**Algorithm 1:** Adaptation Algorithm on Unpaired Data.

```
Input: Unpaired real-world data:  $S_U$ ;  
Paired synthetic data:  $S_P$ ;  
Batch sizes for  $S_U$  and  $S_P$ :  $m$  and  $n$ ;  
Indicator function:  $\mathbf{1}_{S_P}(\cdot)$ .  
1 Load the pretrained models  $P$  and  $D$ .  
2 while not convergent do  
3   Sample unlabeled data  $\{\mathbf{x}_i\}_{i=1}^m$  from  $S_U$ ;  
4   Sample labeled data  $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=m+1}^{m+n}$  from  $S_P$ ;  
5   // Update the primal model  
6   Update  $P$  by minimizing the objective:  
7     
$$\sum_{i=1}^{m+n} \mathbf{1}_{S_P}(\mathbf{x}_i) \mathcal{L}_P(P(\mathbf{x}_i), \mathbf{y}_i) + \lambda \mathcal{L}_D(D(P(\mathbf{x}_i)), \mathbf{x}_i)$$
  
8   // Update the dual model  
9   Update  $D$  by minimizing the objective:  
10    
$$\sum_{i=1}^{m+n} \lambda \mathcal{L}_D(D(P(\mathbf{x}_i)), \mathbf{x}_i)$$
  
11 end
```

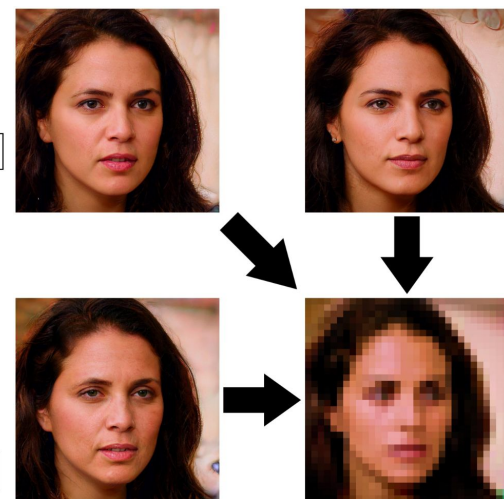
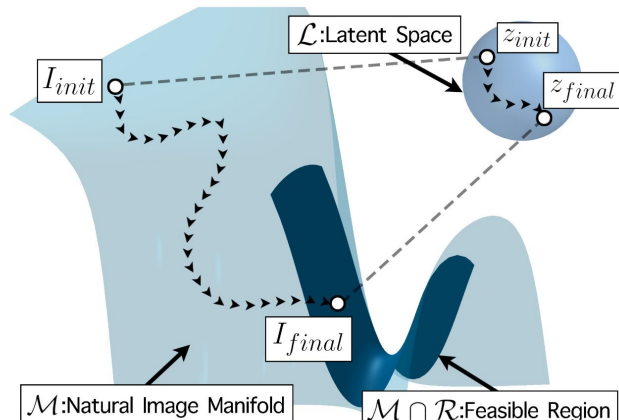
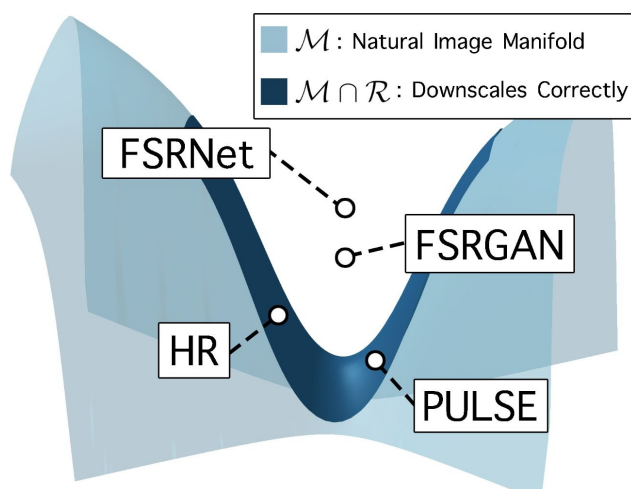
# PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models

会議 : CVPR 2020

著者 : Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, Cynthia Rudin

## PULSE : StyleGANを用いて高倍率でのSRを実現

- 生成したHRをダウンスケールし, 元のLRになる画像を探索する教師なし手法
- HRを生成する潜在空間の事前分布を高次元球面ガウス分布とし,  $\|DS(G(z)) - I_{LR}\|_p \leq \epsilon$  を満たすような $z$ を探索する
  - 潜在空間を限定することで, 生成されるHRは現実的なものに限定される
- Scaling factorは $64 \times$ ,  $32 \times$ ,  $8 \times$ と高倍率を実現
- 顔画像に対して有効性を検証

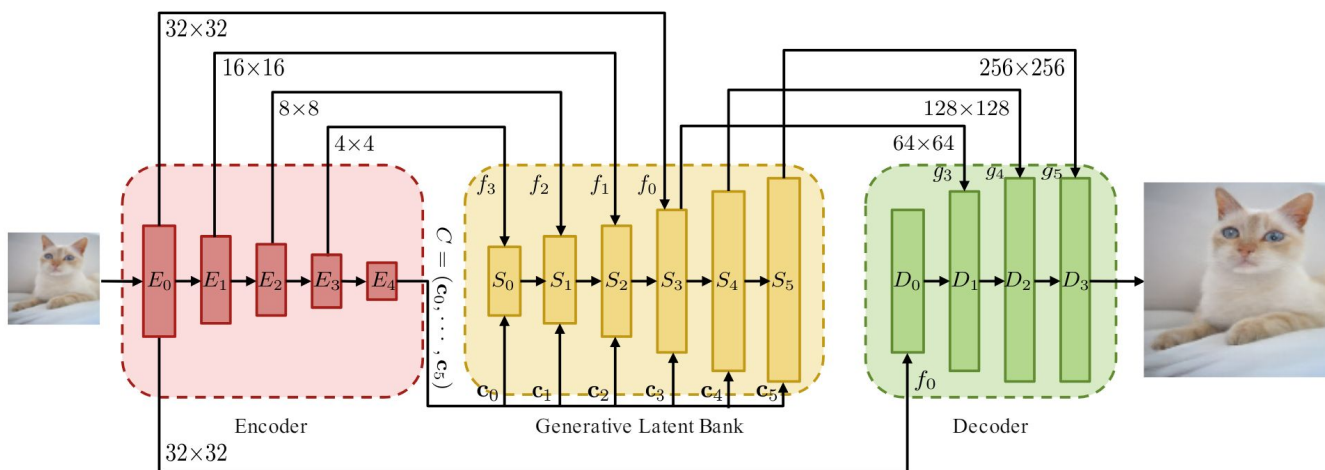


会議 : CVPR 2021

著者 : Kelvin C.K. Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, Chen Change Loy

## GLEAN : StyleGANの利用方法を再考し汎用性向上とコスト削減を達成

- RRDBNetをEncoder( $E_0$ )として特徴を抽出し, 畳み込み層 $E_i$ で解像度事の特徴を取得
- StyleGANに修正を加え, Generative Latent Bankを画像辞書の代わりに導入
  - GANに基づいた辞書は特定の要素や画像に依存せず, サイズと多様性において優れる
  - グローバルマッチングや参照画像/コンポーネントの選択を必要とせず計算効率が高い
- DecoderではEncoderとBankからの特徴を融合
- Scaling factorは $64 \times$  から  $8 \times$



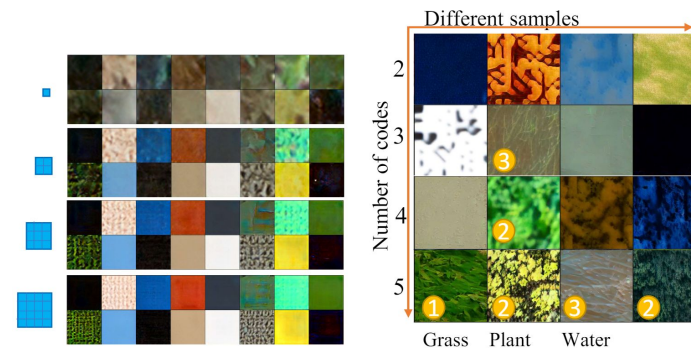
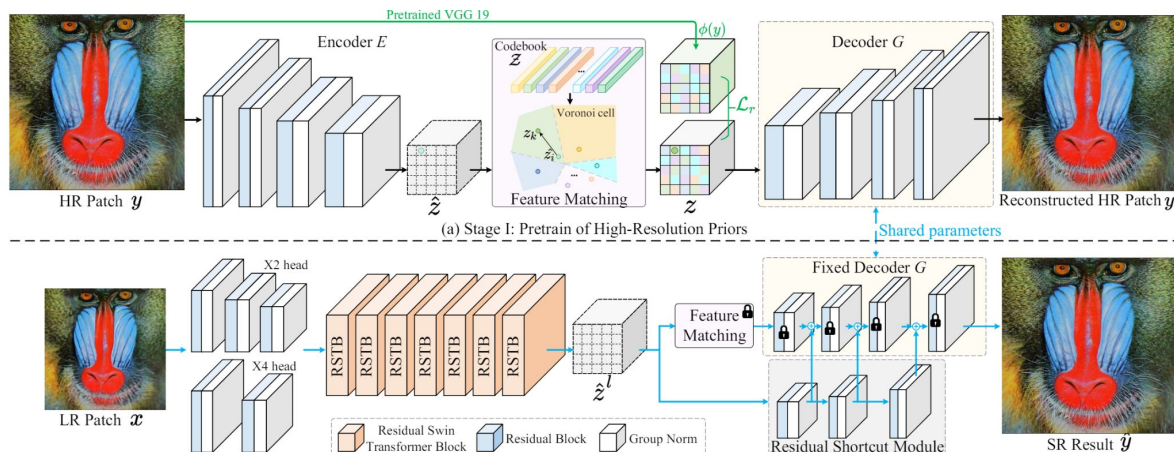


会議 : ACM Multimedia 2022

著者 : Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li

## FeMaSR : VQGANで事前学習された特徴コードブックとLRをマッチング

- LR特徴を高解像度事前分布(HRP)のHR特徴とマッチングする方法でBlindSRを定式化
  - HRPは事前定義された数の特徴ベクトルからなる離散的なコードブックと、対応する事前学習済みデコーダの組み合わせとして定義
- セマンティックに対応したHRPの事前学習のためにセマンティック正則化を導入
- 特徴マッチングのためにHRPへのショートカット接続を持つLRエンコーダを導入



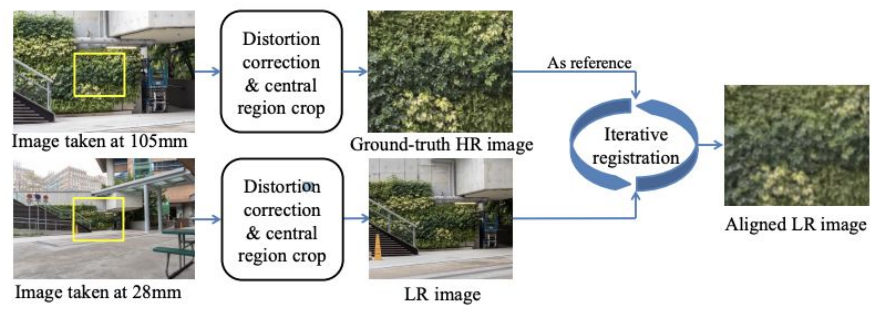
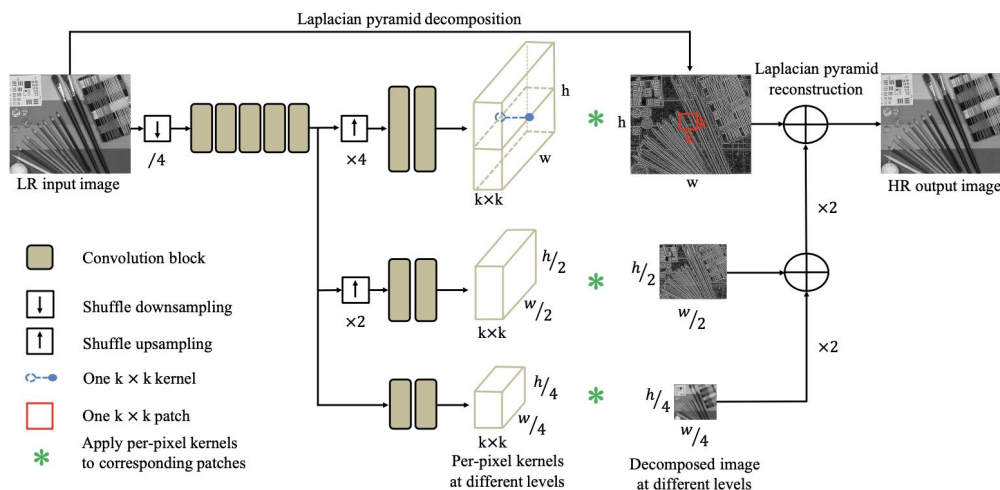


会議 : ICCV 2019

著者 : Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang

## RealSR: 実世界の劣化に対応したデータセットとネットワークを提案

- 一眼カメラでHRとLRのペアを撮影し, 実世界のSRモデルの学習/評価データセットを提案
  - HRとLRのペアは焦点距離を変えて, 別々に撮影後, 位置合わせして作成
- 実世界のSRにおいて, 従来のネットワークより計算コストを削減しつつ性能を向上させた Laplacian pyramid based kernel prediction network (LP-KPN) を提案
  - 様々な劣化に対する効率的な学習のため, 各ピクセルごとに個別のカーネルを学習
  - 画像をラプラシアンピラミッドに分解し, より効率的に豊富な近傍情報を利用

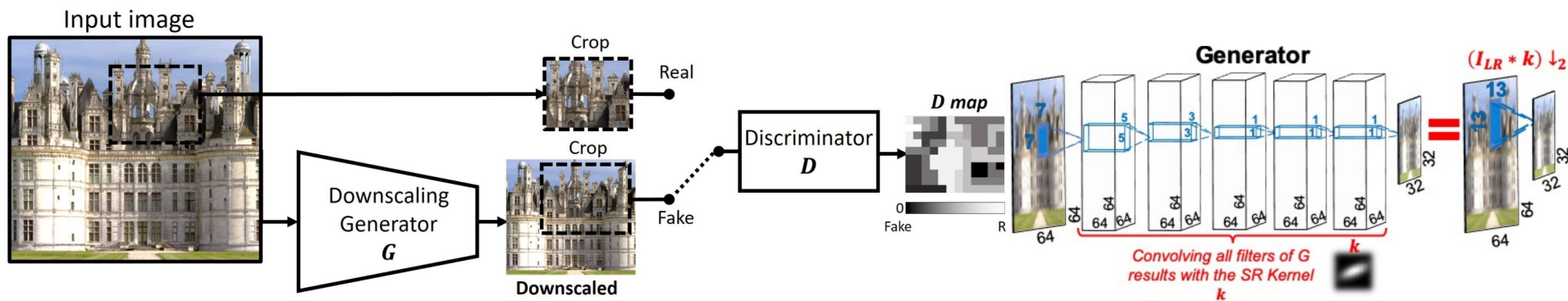


会議 : NeurIPS 2019

著者 : Sefi Bell-Kligler, Assaf Shocher, and Michal Irani

## KernelGAN: パッチの分布を最も保存するSRカーネルを推定するKernelGANを提案

- LRテスト画像のみを用いて完全な教師なしでGeneratorをトレーニング
  - LR画像のパッチ分布にできるだけ近いダウンスケーリングされた画像を生成
- Discriminatorは各ピクセルに対して、その周囲のパッチが元のパッチ分布からどれだけの確率で描画されるかを示すヒートマップを出力
  - LRのパッチ分布を学習し、Generatorで生成したパッチを区別する
- 活性化関数を用いない線形多層NNをGeneratorとして使用
  - 表現能力は単層のConvolution layerと同等だが、非凸な最適化に有効
- SRカーネルは画像特有であるだけでなく、異なるスケールのカーネルを導出可能

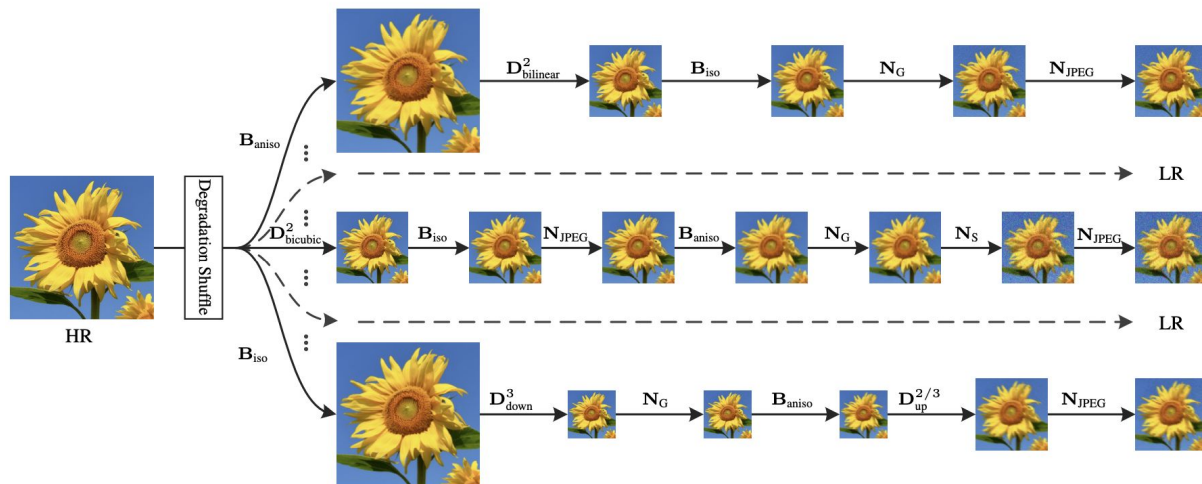
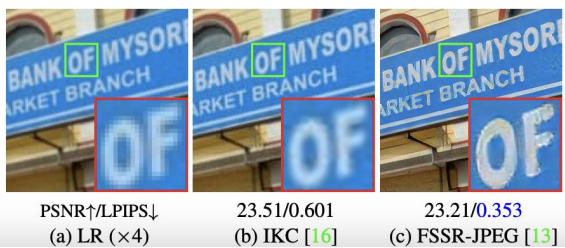


会議 : ICCV 2021

著者 : Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte

## BSRGAN: 実世界の劣化過程を考慮した学習/評価パイプラインを提案

- ブラー, ダウンサンプリング, ノイズに対するより複雑な劣化を考慮した実用的なSR劣化モデルを設計
- より詳細な評価を行うため, DIV2K4DとRealSRSetの2つの評価データセットを導入, 提案した劣化モデルが優れていることを示した
  - PSNR/LPIPSの評価で既存手法を上回る性能
  - non-reference型の評価指標であるNIQE, NRQM, PIは定性評価と必ずしも一致しない

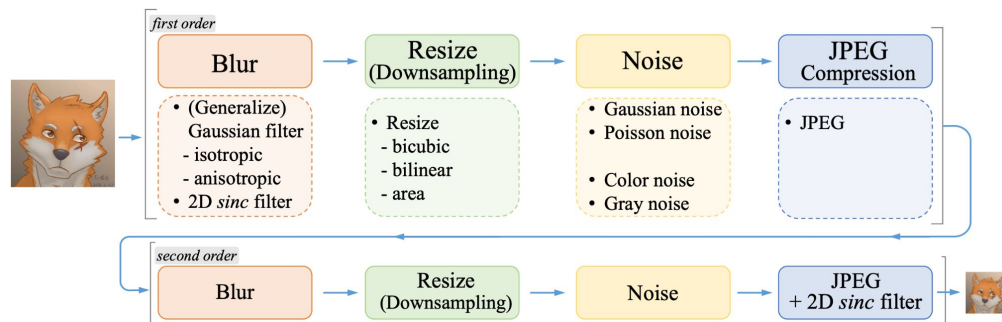


会議 : ICCV 2021 Workshop

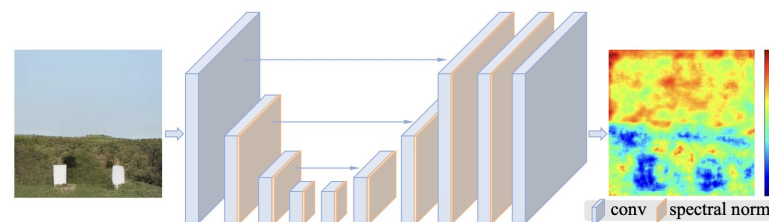
著者 : Xintao Wang, Liangbin Xie, Chao Dong, Ying Shan

## Real-ESRGAN: ESRGANを実世界の劣化に対応させ実用性を向上

- ESRGANのHR-LRペア作成時の劣化処理およびDiscriminatorを見直すことで実世界の劣化に対応
  - 実世界の劣化を考慮した劣化プロセスを提案
    - 劣化処理をbicubic補間から実世界の劣化を考慮した処理に変更
    - さらに、従来の実世界に対応させるための劣化過程はインターネット上での拡散による劣化を考慮していなかったが、複数回繰り返すことで対応
    - オーバーシュートやリングングに対応するため、sinc関数による劣化処理を追加
  - DiscriminatorをVGGベースからU-Net構造へ変更
    - ピクセル単位の情報を考慮した学習が可能となり、識別能力が向上
    - U-Net構造と複雑な劣化は学習を不安定にするため、spectral normalization regularizationで安定化



実世界の劣化をモデル化するためのn次劣化プロセス  
(論文中ではn=2を採用)



Discriminator構造

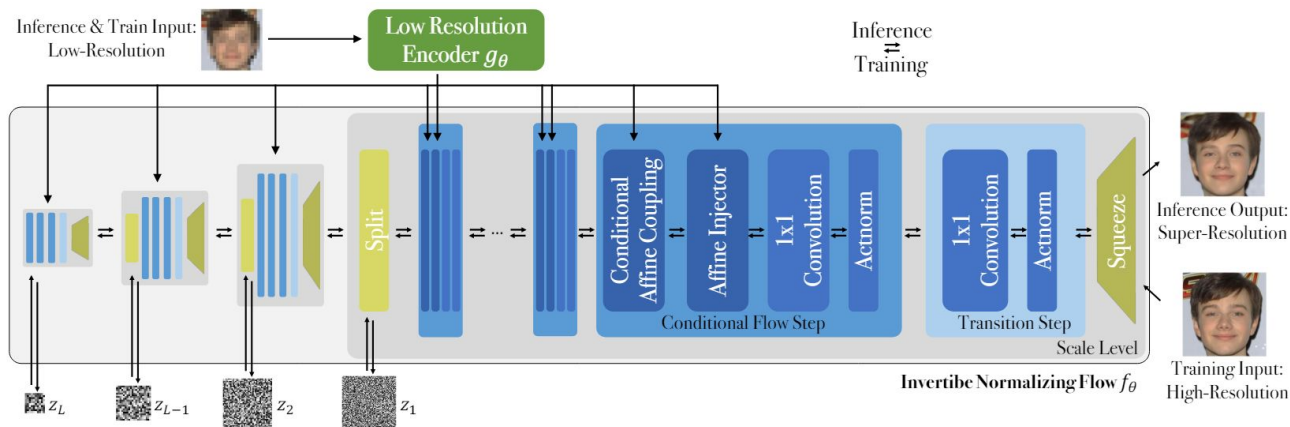


会議 : ECCV 2020

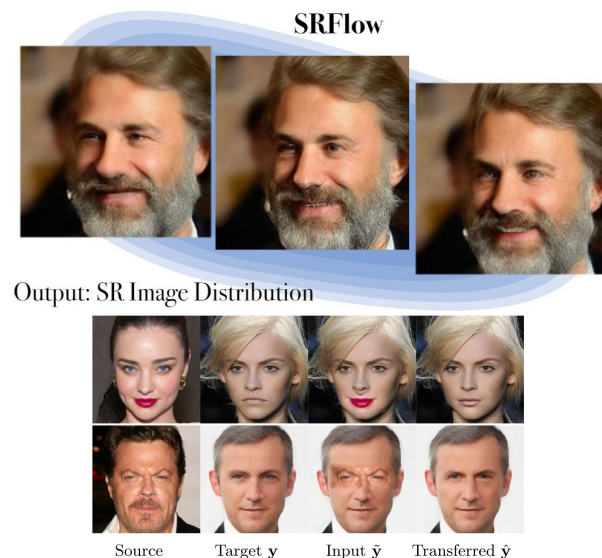
著者 : Andreas Lugmayr, Martin Danelljan, Luc Van Gool, Radu Timofte

## SRFlow : Flowを超解像に導入

- 単一のHR画像を予測するのではなくHR画像の分布をモデル化するため柔軟な予測が可能
  - 学習時: HRを可逆計算可能な層に複数回通すによりHRの潜在分布 $p(z)$ を獲得
  - 推論時: LRと $z$ から訓練時の逆の演算によりHRを復元
- 可逆な構造であるため, 潜在空間と画像空間の両方で柔軟な操作が可能
  - 様々なアプリケーションや画像操作タスクに適用可能
- モード崩壊がなく, 単一の損失関数で安定した学習が可能



可逆な構造を持ち, 学習( $\leftarrow$ )と復元( $\rightarrow$ )を行うことができる





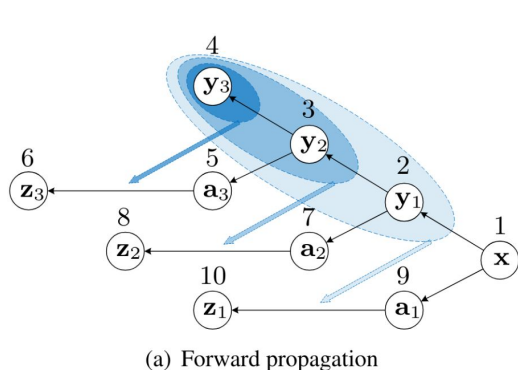
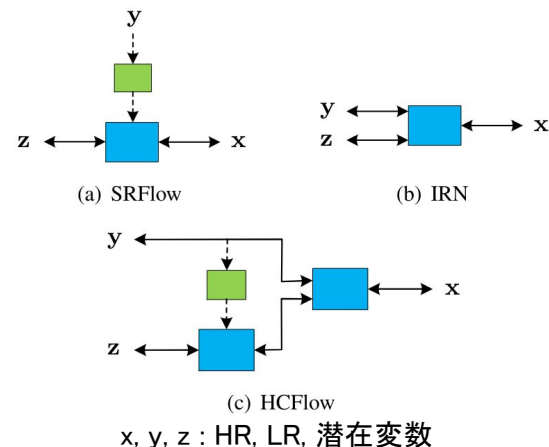
# Hierarchical Conditional Flow: A Unified Framework for Image Super-Resolution and Image Rescaling

会議 : ICCV 2020

著者 : Jingyun Liang, Andreas Lugmayr, Kai Zhang, Martin Danelljan, Luc Van Gool, Radu Timofte

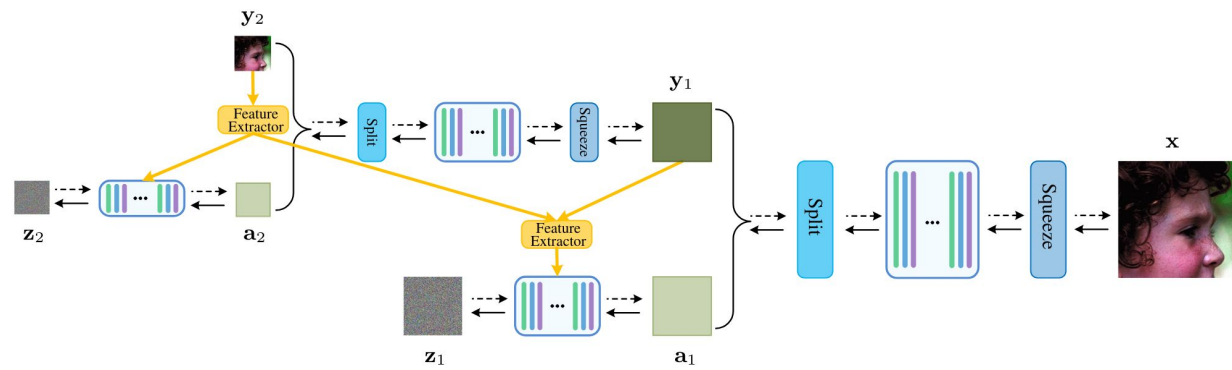
## HCFlow: LRとHRの可逆性をより追求したFlowモデル

- LR画像と残りの高周波成分の分布を同時にモデル化することで, HR画像とLR画像のペアの双射 (bijective) マッピングを学習
  - 順伝搬時にHRをLRと潜在変数に分割することを学習する
  - LRと潜在変数からHRを復元
- HRは**階層的(L段)**にLRと高周波成分に分割され, は  $z_l$  に条件付けられている
- **対数尤度損失に加え**, ピクセル損失, 知覚損失, GAN損失も学習時に組み合わせる



(a) Forward propagation

x, y, a : HR, LR, 高周波成分

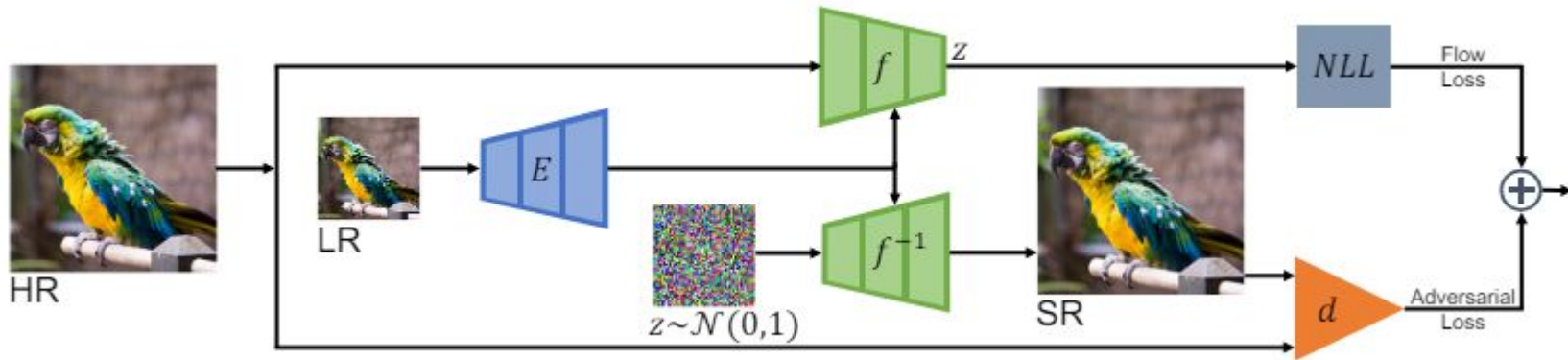


会議 : WACV 2022

著者 : Andreas Lugmayr, Martin Danelljan, Fisher Yu, Luc Van Gool, Radu Timofte

## AdFlow: L1損失の代替としてFlowを使用

- 従来のL1損失と敵対的損失は競合しており, LRの一貫性に悪影響を与えることから, L1損失を確率的な視点から再考し, 一層の条件付きflowとして表現
  - “不確実な”ピクセルに対して大きなスケール値  $b(x)$  を予測するよう促し, HR画像のより柔軟な分布を予測
  - 多層のflowを用いることでピクセル同士の相関を捉えた損失を提案
  - 従来のL1損失に比べ, 敵対的損失との相性が良い



$$-\log p(y|x; \theta) \propto \left\| \frac{y - g(x)}{b(x)} \right\|_1 + \sum_{ijc} \log b(x)_{ijc}.$$

会議 : TPAMI 2022

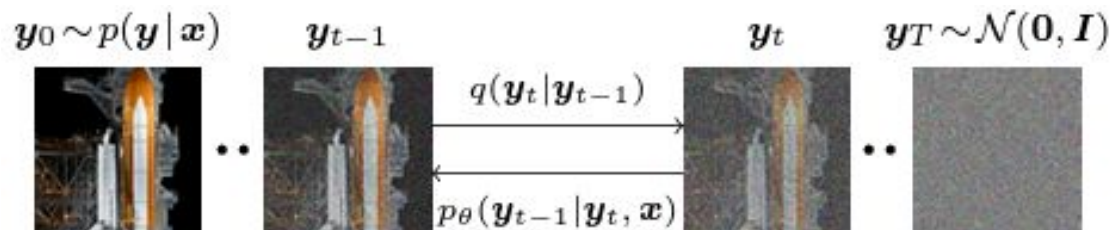
著者 : Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, Mohammad Norouzi

## SR3: GANやFlowベースの既存手法を上回る性能を達成

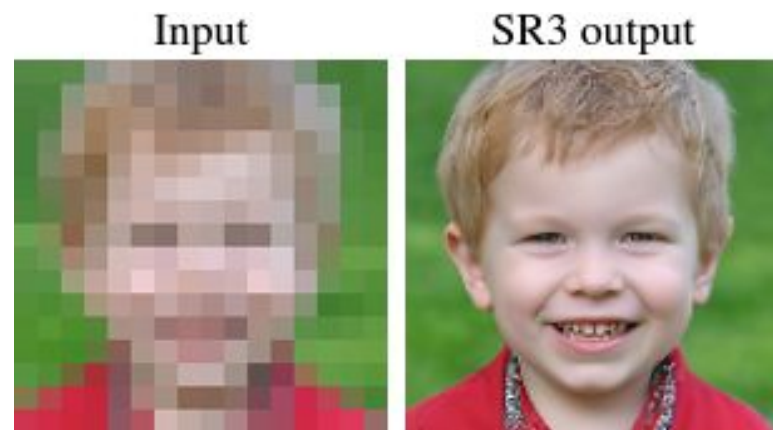
- SR3: Googleが2023年に特許出願済みの技術

- 「機械学習モデルに基づく反復改良による画像強化」

- ノイズ除去拡散確率モデルを条件付き画像生成に適応させ、確率的反復ノイズ除去プロセスを通じて超解像度を実行



- 顔画像の8倍up samplingタスクで、人間の評価者が「生成画像か否か」を判別する評価において、評価者を50%騙すことができた (生成画像に対して、これは生成画像ではないという判断を下した率)



会議 : Neurocomputing 2022

著者 : Haoying Li, Yifan Yang, Meng Chang, Huajun Feng, Zihai Xu, Qi Li, Yueting Chen

## SRDiff: Diffusionの利用, 多様な表現を持った復元画像を生成可能

- DiffusionベースのモデルでSingle Image Super-Resolution (SISR)を実装した最初の論文
  - 過剰な平滑化, モード崩壊, 巨大なモデルのスペース問題 を解決できる
    - large model footprint: モデルがメモリの的に巨大なスペースを必要とする問題
  - 多様な表現, 属性を高解像度で表現できる
- Small footprintで安定的／効率的に学習可能
  - 単一の損失関数を利用するため, GANよりも安定的に学習可能
  - アーキテクチャ上の制約がないため, Flowベースよりもfootprintが小さい

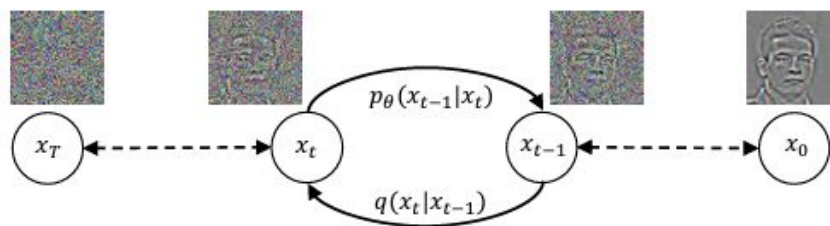
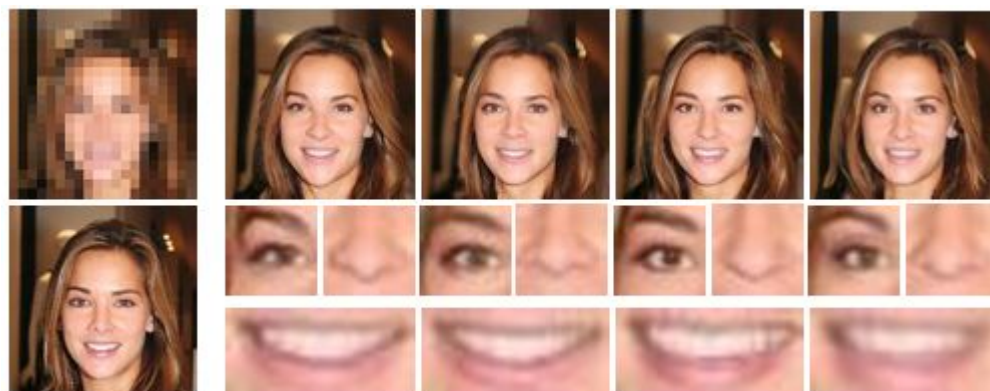


Figure 2: Overview of two processes in SRDiff. The diffusion process is from right to left and the reverse process is from left to right.  $\theta$  in  $p_\theta$  denotes the learnable components including conditional noise predictor and low-resolution encoder in SRDiff.





会議 : ICCV 2023

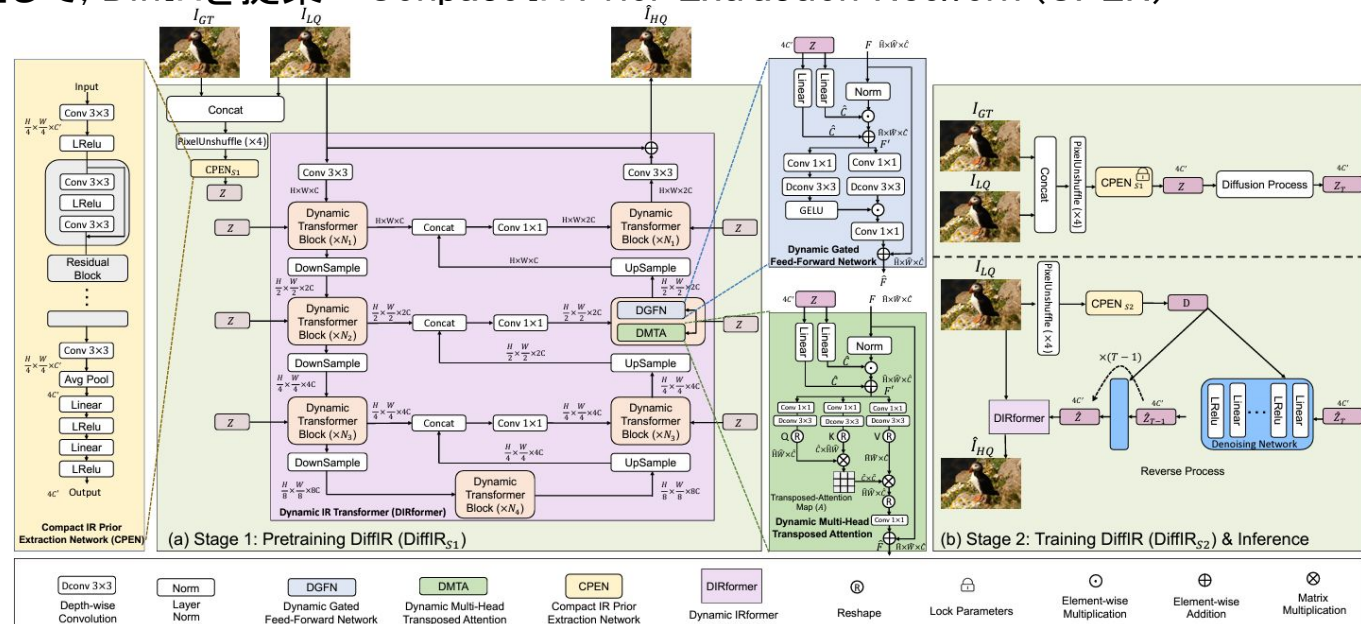
著者 : Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, Luc Van Gool

## DiffIR: 従来のDiffusionモデルよりも少ないiterationで正確な推定・安定的で現実的な出力結果を実現

- Diffusion Model (DM)は、画像復元 (Image restoration; IR) に応用しようとすると大規模なiterationで学習を回す必要があった。

- IRのためのDMとして, DiffIRを提案 → Compact IR Prior Extraction Network (CPEN)

- Step 1  
CPENの事前学習.  
コンパクトなIRの事前表現を獲得
- Step 2  
DMの直接的な学習.
- DiffIRはコンパクトな特徴ベクトルのみで、安定的で現実的な復元画像を生成可能



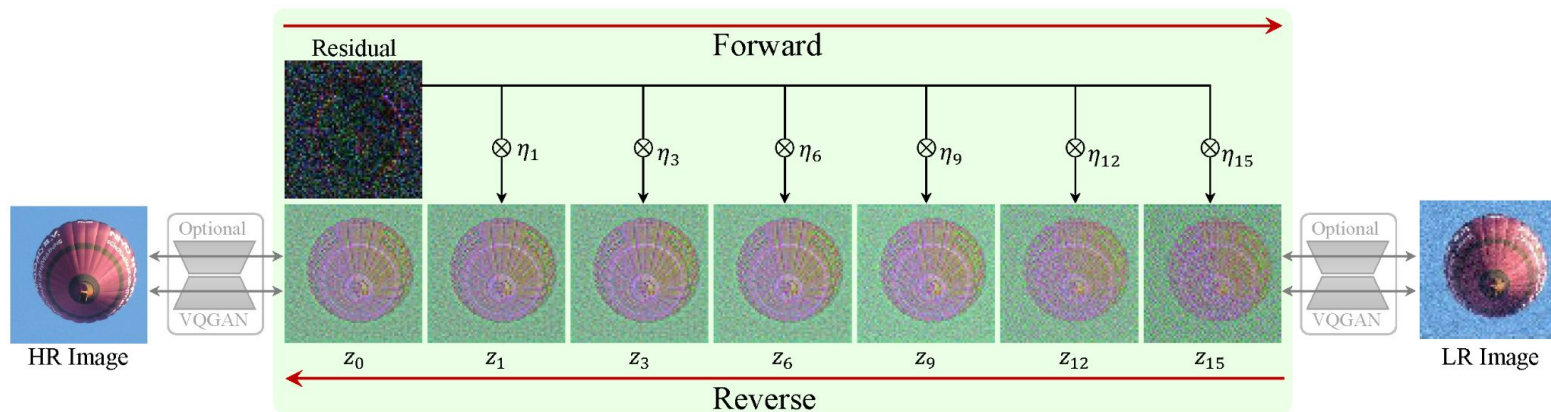
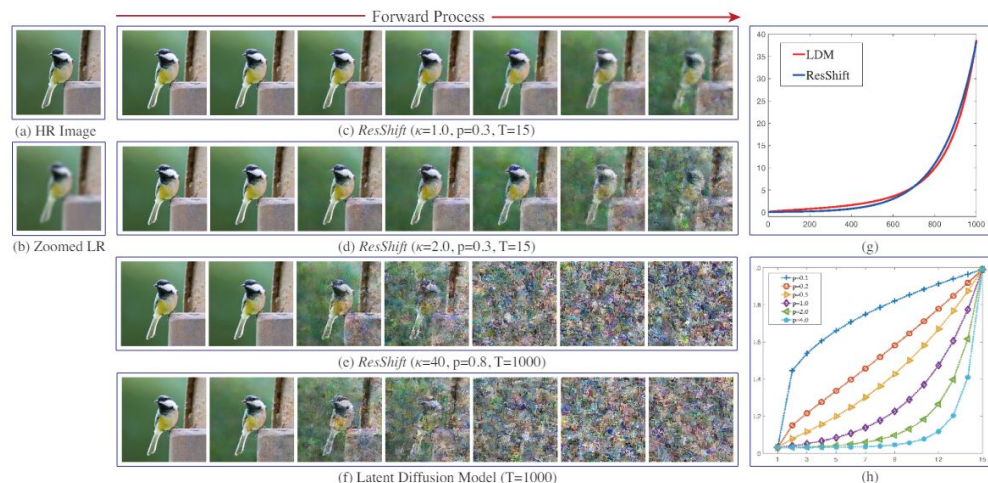


会議 : NeurIPS 2023

著者 : Zongsheng Yue, Jianyi Wang, Chen Change Loy

## ResShift: 拡散モデルのサンプリングステップを15に削減!

- 品質を犠牲にせずに拡散モデルのサンプリングステップを短縮
- LRとHRの残差をマルコフ連鎖を通して徐々にシフトしていくことで復元
- 柔軟なノイズスケジュールを行うことでSRの忠実性と現実性のトレードオフを実現



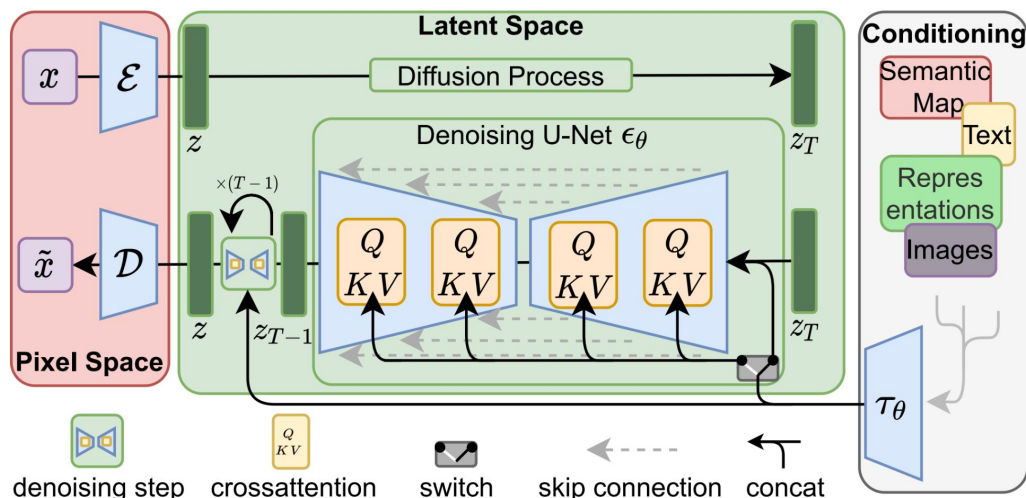
会議 : CVPR 2022

著者 : Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Bjorn Ommer

## LDM: 潜在空間での学習と多様な条件付け機構により, 低コストかつ高品質な画像生成が拡散モデルで可能に

- 潜在空間での拡散モデルの学習
  - 生成画像の品質を保ちつつ, 計算効率を大幅に向上
- 多様な条件付け機構
  - クロスアテンションを用いて, テキストをはじめとした多様な入力条件に応じた画像生成が可能
  - 超解像の場合は低解像度画像で条件づけ

DDPMベースの超解像モデル SR3をFID, ISで上回る結果

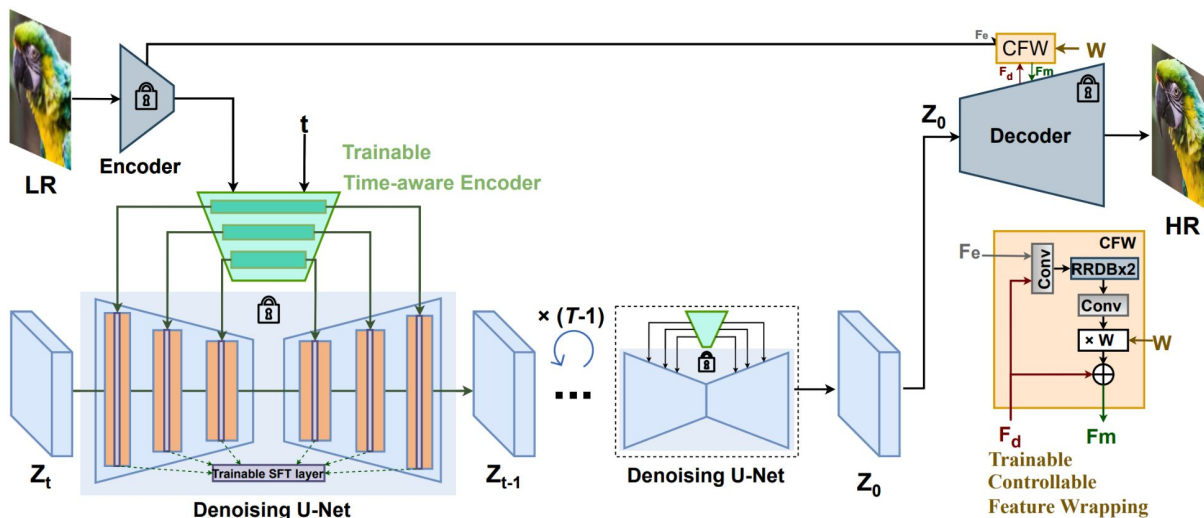


会議 : arXiv 2023

著者 : Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C.K. Chan ,and Chen Change Loy

## StableSR: 事前学習済み拡散モデルにpriorを追加し超解像向けにチューニング

- time-aware encoder
  - 事前学習済みの拡散モデルに追加することで, 性能低下させずに学習時間を短縮
- controllable feature wapping module
  - ユーザーが推論プロセス中にスカラー値を調整し, 品質と忠実度のバランスを自由に調整可能にすることで拡散モデルでの推論結果のランダム性に対処
- progressive aggregation sampling strategy
  - 事前学習した拡散モデルの固定サイズの制約を克服するために, 任意の解像度に適応





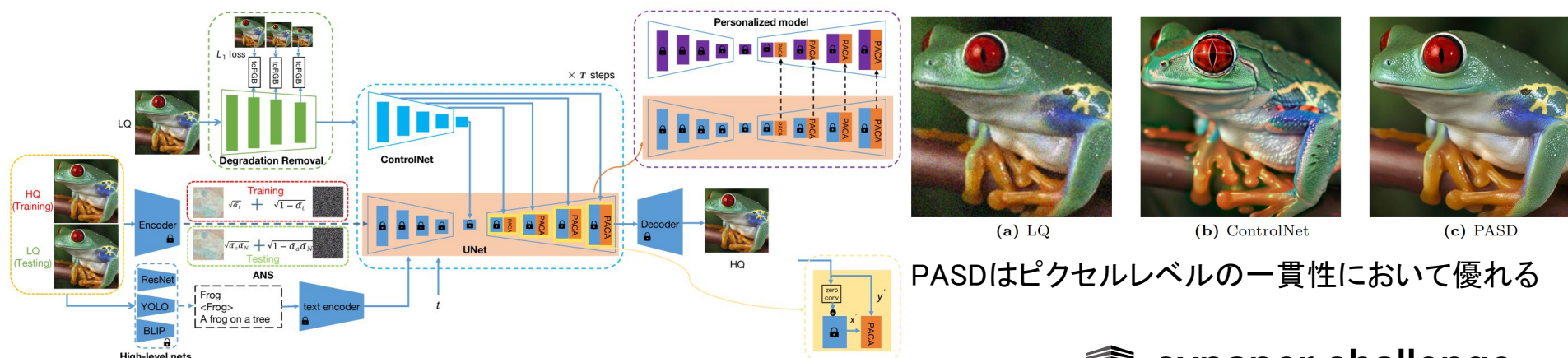
# Pixel-Aware Stable Diffusion for Realistic Image Super-Resolution and Personalized Stylization

会議 : arXiv 2023

著者 : Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, Lei Zhang

## PASD:ピクセルレベルの忠実性を強化

- pixel-awareクロスアテンション(PACA)の導入によりピクセルレベルの局所特徴を認識
  - 拡散プロセスにピクセルを考慮した条件付き制御を導入
- 劣化除去モジュールにより劣化の影響を受けない特徴を抽出
  - マルチスケールの特徴マップを抽出し, 各スケールでHRを近似
- 調整可能なノイズスケジューリング(ANS)の導入
  - LRの影響を調節することにより, 復元時の強い劣化による悪影響を軽減
- 事前学習済みResNet, YOLO, BLIPによりLRの高レベルの情報を抽出
- ベースモデルを変更することで, 複数の高度な画像処理が可能



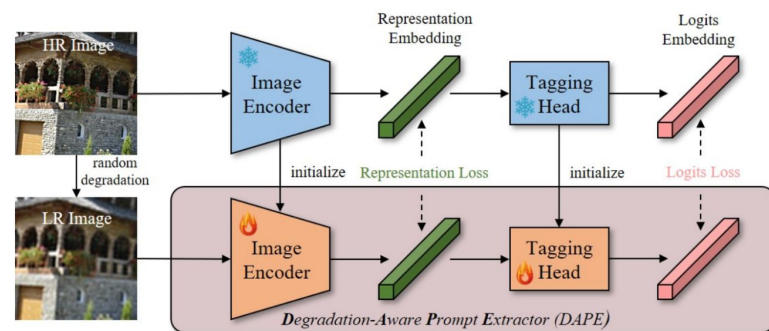
PASDはピクセルレベルの一貫性において優れる

会議 : CVPR 2024

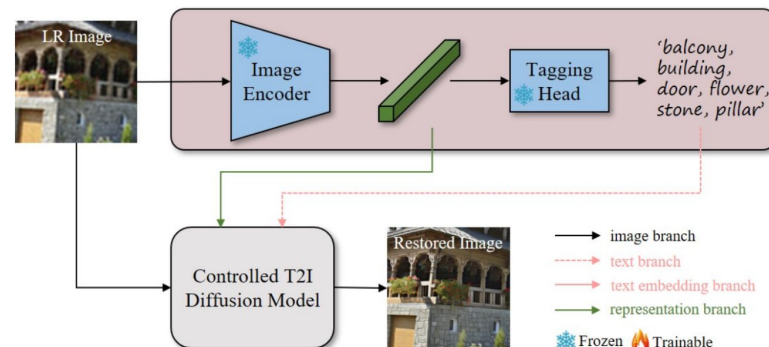
著者 : Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, Lei Zhang

## SeeSR: 意味的プロンプトを用いて意味的忠実性を向上

- 劣化による構造情報の損失によって生じるsemantic errorの改善を目指した論文
- 劣化認識プロンプト抽出器としてDegradation-aware prompt extractor (DAPE)を提案
  - 劣化画像とGT画像から同じ意味合いを持つプロンプトを出力するように学習
    - ソフトプロンプト(特徴埋め込みベクトル): 画像の全体的な意味情報や文脈を補完するための表現情報
    - ハードプロンプト(カテゴリラベル): T2Iモデルの特定のオブジェクト認識を向上させるための情報
- DAPEから出力したプロンプトを用いることで現実世界の複雑な劣化に対しても, T2Iモデルが正確かつ詳細な画像を再現可能に



(a) Degradation-aware prompt extractor



(b) Real-ISR with DAPE



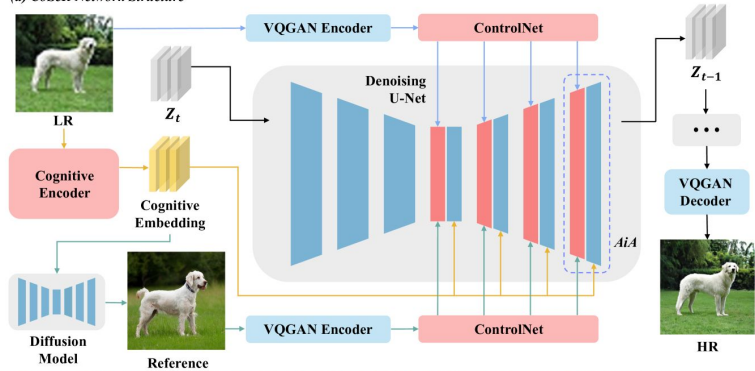
会議 : CVPR 2024

著者 : Haoze Sun, Wenbo Li, Jianzhuang Liu, Haoyu Chen, Renjing Pei, Xueyi Zou, Youliang Yan, Yujiu Yang

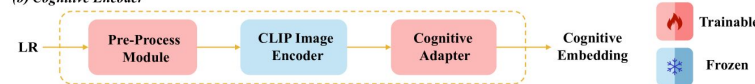
## CoSeR: LR画像を認知し, 生成モデルの知識を利用して画像を復元

- 低解像度画像からDiffusionモデルを利用してReference画像を生成
- 認知→reference画像を生成→復元という, より人間による画像復元に近いプロセスを再現
  - 事前学習済みCLIPの画像エンコーダにより画像の意味的特徴を抽出
  - 画像特徴からマルチトークン埋め込みを抽出するアダプターを導入
- 認知埋め込み, 参照画像, LRからの情報を統合するAll-in-Attentionを提案
- 人間の知覚ベースの評価指標において高い精度を達成

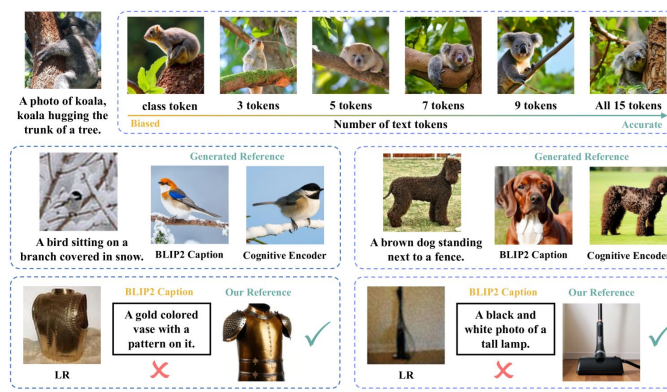
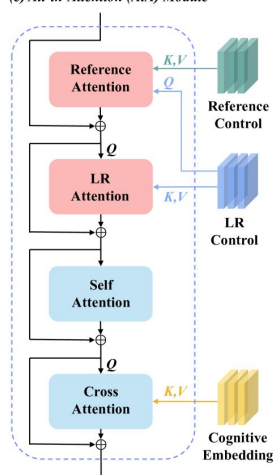
(a) CoSeR Network Structure



(b) Cognitive Encoder



(c) All-in-Attention (AiA) Module



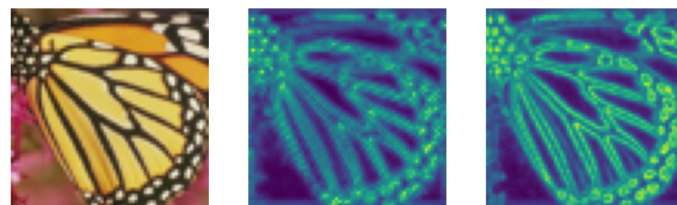
CLS以外のトークンも利用することで  
意味を忠実に表現

会議 : CVPR 2024

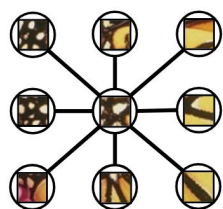
著者 : Yuchuan Tian, Hanting Chen, Chao Xu, Yunhe Wang

## IPG : CNNやAttentionの代わりに柔軟なグラフを利用

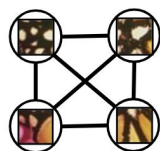
- 画像のピクセル同士の連関をGNNで柔軟に表現
  - SRタスクでは高周波部分が重要であり, ピクセルの重要度が一定ではない
  - 画像ノードの重要度を測定するためにディテールを考慮した尺度を提案し, 高周波のノードに大きなノード次数をつける
    - pixelの重要度はdownsampling→upsamplingで変化する量に基づき決定
  - パッチではなくピクセルをグラフノードとする
  - 局所および大域の両方を考慮した効率的なサンプリングを提案
- **CVPR 2024 Best Student Paper Runners-Up !**



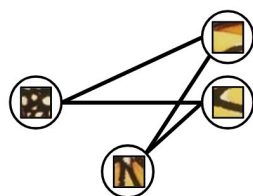
提案された尺度に基づくdetail richなピクセル



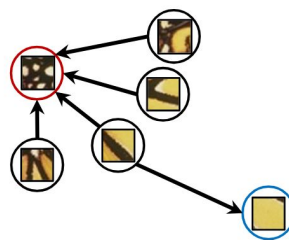
Convolution



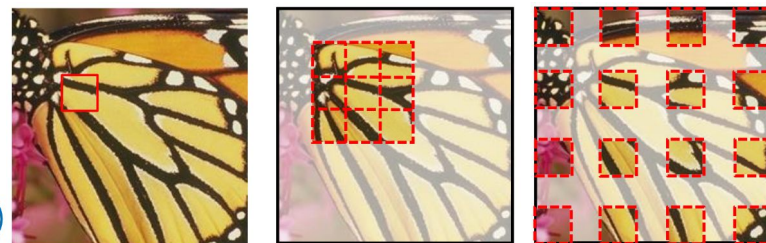
Self-Attention



KNN Graph



Ours



効率的なサンプリング戦略

メタサーベイ&最新トレンド

# メタサーベイ(1/2)

---

## 超解像のメタ話

- ❑ デファクトスタンダードなSRモデルは他分野モデルのアーキテクチャを取り入れている
  - ❑ VGG Network → VDSR
  - ❑ ResNet → SRResNet
  - ❑ SRDenseNet → DenseNet
  - ❑ IPT → Transformer
  - ❑ SwinIR → Swin Transformer
  - ❑ MambaSR → Vision Mamba
- ❑ SRモデルを提案するには他分野のモデルもサーベイしておく必要がある？

# メタサーベイ(2/2)

## 超解像のメタ話

- Workshop論文のモデルがデファクトスタンダードとなるケースがいくつか見られる(データセットでも)

- モデル: EDSR(6,766), ESRGAN(4,257), SwinIR(2,439), Real-ESRGAN(921),

- データセット: DIV2K(3,202), Flickr2K(1,701)

( )は2024/6/25時点でのgoogle scholar被引用数

- 画像復元分野の主要なWorkshopは以下の二つ

-  **NTIRE** (New Trends in Image Restoration and Enhancement workshop)  
2016年からCVPRで毎年開催

-  **AIM** (Advances in Image Manipulation workshop)  
2019年からICCVもしくはECCVで開催

- 上記のモデルは全てAward獲得

- EDSR → ”NTIRE2017 1st Place Award”

- ESRGAN → ”Champion PIRM Challenge on Perceptual” Super-Resolution  
(ECCV2018のみ開催)

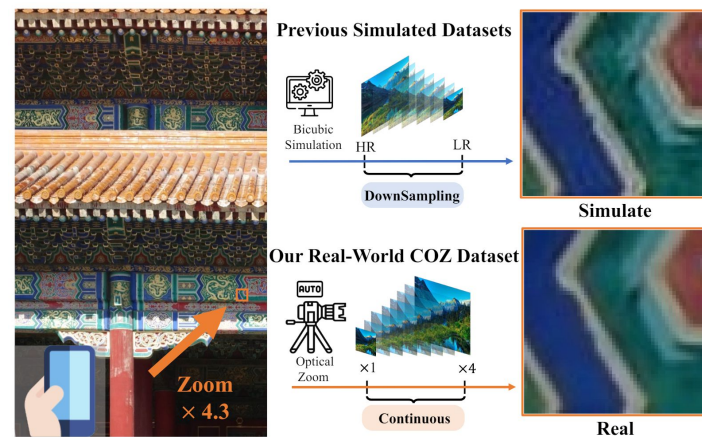
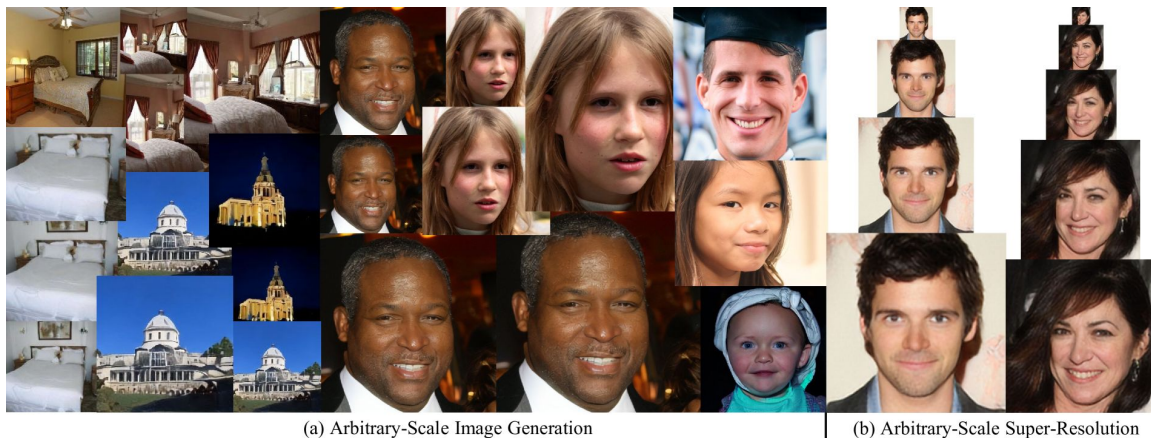
- SwinIR → ”AIM2021 Best Paper Award”

- Real-ESRGAN → ”AIM2021 Honorary Nomination Paper Award”



# 最新トレンド(1/5)

- ❑ 従来の2倍/4倍といった固定のスケールではなく、**任意スケール**に対応した超解像が注目されていると感じる
- ❑ CVPR2024でもベンチマークの提案があった: COZデータセット



COZ データセット論文より引用

<https://arxiv.org/pdf/2403.10255v1>より引用

# 最新トレンド(2/5)

## 超解像のメタ話

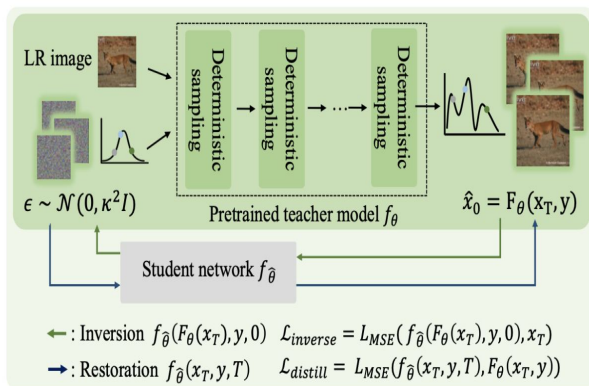
### ❑ Diffusion ベースの手法の増加

#### ❑ ステップ数の削減

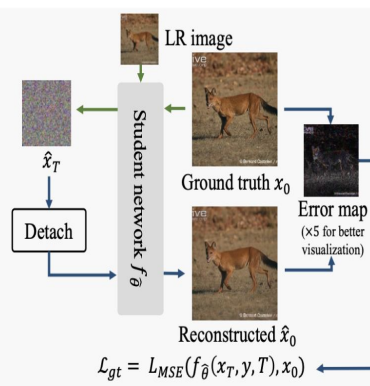
#### ❑ LRとの一貫性の保持

#### ❑ 任意スケールの超解像

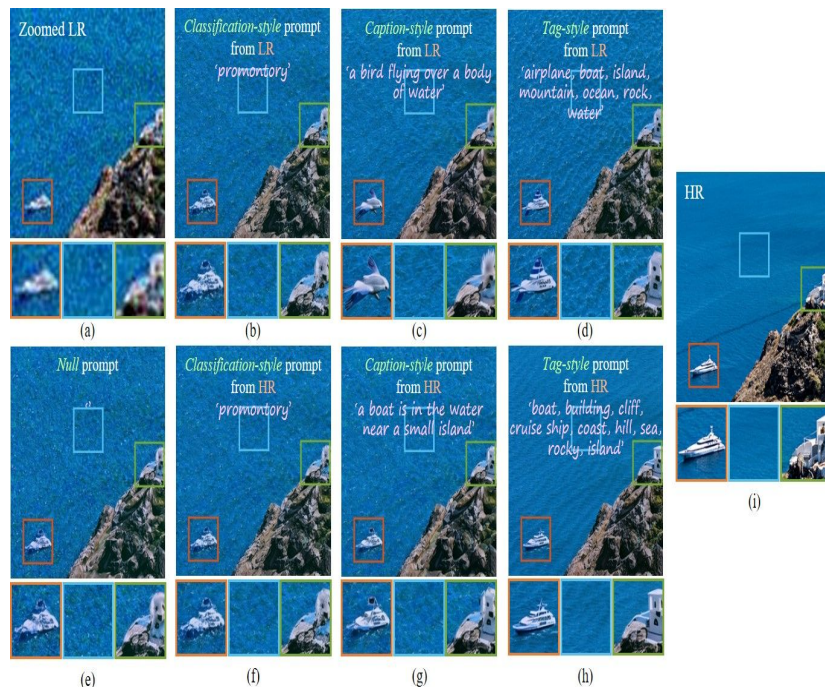
### ❑ LR画像に映っていないものを生成させない工夫が重要



(a) One-step bi-directional distillation



(b) Consistency preserving loss

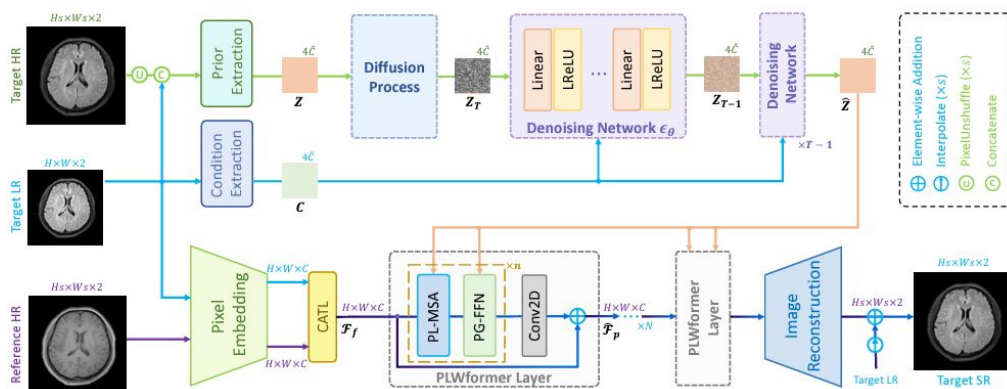


[https://openaccess.thecvf.com/content/CVPR2024/papers/Wu\\_SeeSR\\_Towards\\_Semantics-Aware\\_Real-World\\_Image\\_Super-Resolution\\_CVPR\\_2024\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2024/papers/Wu_SeeSR_Towards_Semantics-Aware_Real-World_Image_Super-Resolution_CVPR_2024_paper.pdf) より引用

<https://github.com/wyf0912/SinSR> より引用

# 最新トレンド(3/5)

- Diffusionモデルによる生成画像を参考にする手法が増加傾向
- 長所: 意味的に正しく, かつ写真のようにリアルなディテールを復元できる



- テキストでガイドを入れる手法も最近人気

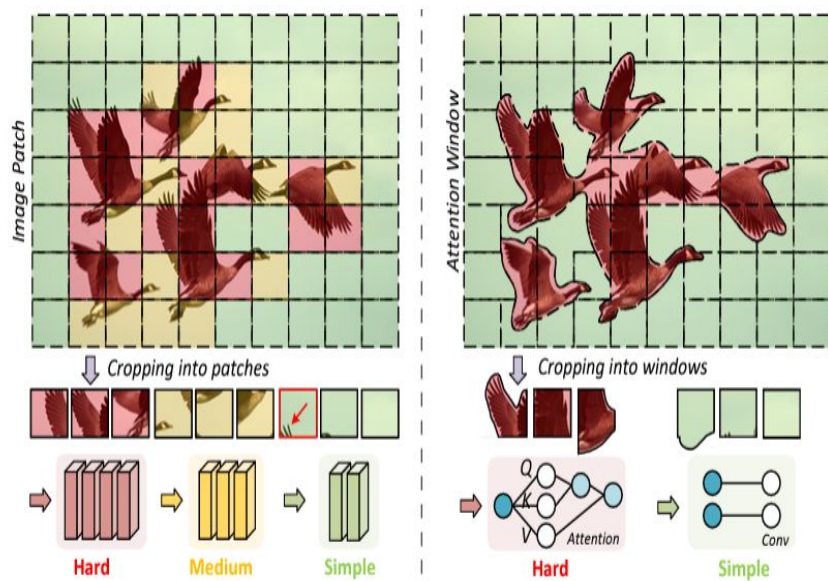


LR input DPS [12] using diffusion model trained on Imagenet Ours  
'A statue of Walt Disney holding Mickey Mouse hands is showing in front of Cinderellas castle.'

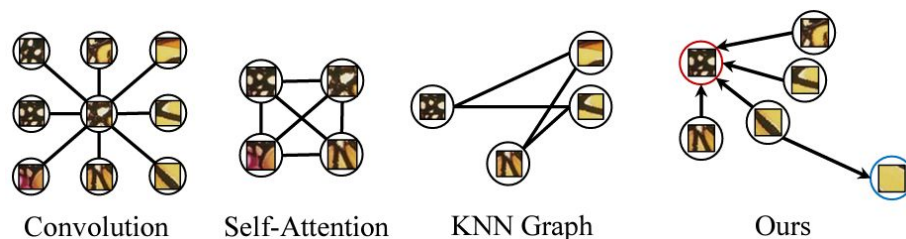


# 最新トレンド(4/5)

- ❑ 一枚画像中で異なる復元の難易度に着目する話が多くなってきているのかも？
- ❑ 超解像では、復元が難しい領域と簡単な領域が一枚画像中に混在している。
  - ❑ パッチごとにHard/Medium/Simpleで難易度分けするClassSRを代表として、一枚画像中で難易度分け → 割り当てるパラメータ量を変化させることにより超解像処理の計算量を削減する試みが増えてきている



[CAMixerSR: Only Details Need More "Attention"](#)



[Image Processing GNN: Breaking Rigidity in Super-Resolution](#)



Figure 4. Visualization of adaptive bit-mapping of AdaBM. Results are from EDSR-AdaBM (w8 a6MP).

[AdaBM: On-the-Fly Adaptive Bit Mapping for Image Super-Resolution](#)

# 最新トレンド(5/5)

## CVPR Workshop : [NTIRE 2024](#)

- ❑ 新しいトレンドは**Mamba**かも？
- ❑ MambaSRがNTIRE 2024 Challenge on Image Super-Resolution (×4)にて最高性能を達成！
- ❑ Mamba: 状態空間モデルの利用
  - ❑ MambaSRではState Space Transformer (SSFormer)を導入
  - ❑ HATをベースとして, Hybrid Attention BlocksをSSFormerによって置換

Team	Rank	PSNR (primary)	SSIM
XiaomiMM	1	31.94	0.8778
SUPSR	2	31.41	0.8711
UCAS-SCST	3	31.28	0.8666
SYSU-SR	4	31.19	0.8660
Jasmine	5	31.18	0.8665
ACVLAB	6	31.18	0.8655
mandalinadagi	7	31.13	0.8648
SKDADDYS	8	31.11	0.8643
KLETech-CEVI	9	31.03	0.8633
SVNIT-NTNU	10	31.03	0.8633
ResoRevolution	11	31.01	0.8647
BetterSR	12	30.97	0.8621
Contrast	13	30.69	0.8563
BFU-SR	14	30.55	0.8560
SCU-VIP-LAB	15	29.78	0.8506
Nudter	16	30.17	0.8446
JNU-620	17	30.43	0.8426
LVGroup-HFUT	18	29.98	0.8380
Uniud	19	29.97	0.8440
SVNIT-NTNU-1	20	29.34	0.8199

