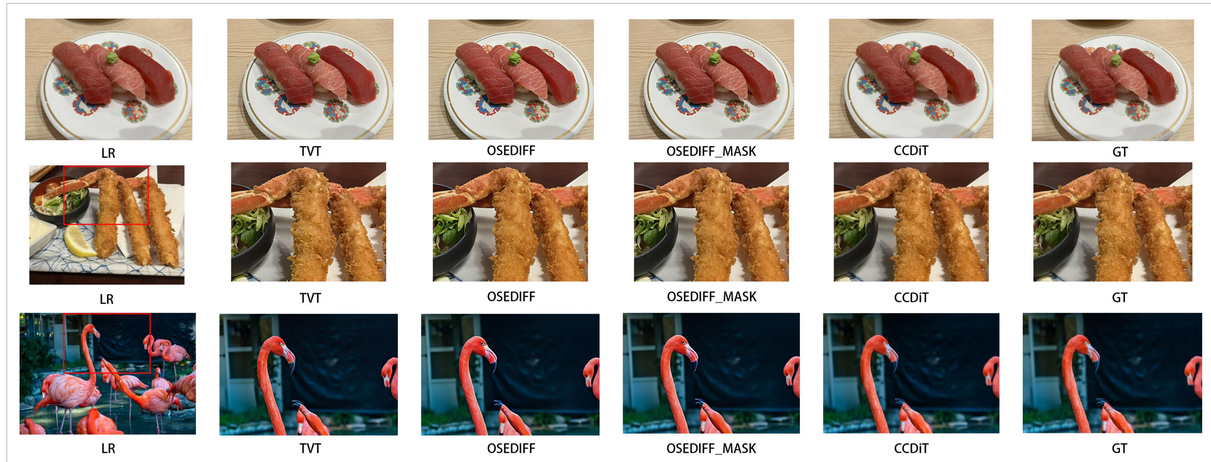
IS3-176

CCDiT：色一貫性を考慮した 単一ステップ拡散 Transformer による画像超解像

Yitian MA, Longhao SUN, Peilin XIONG, Keiji YANAI
電気通信大学

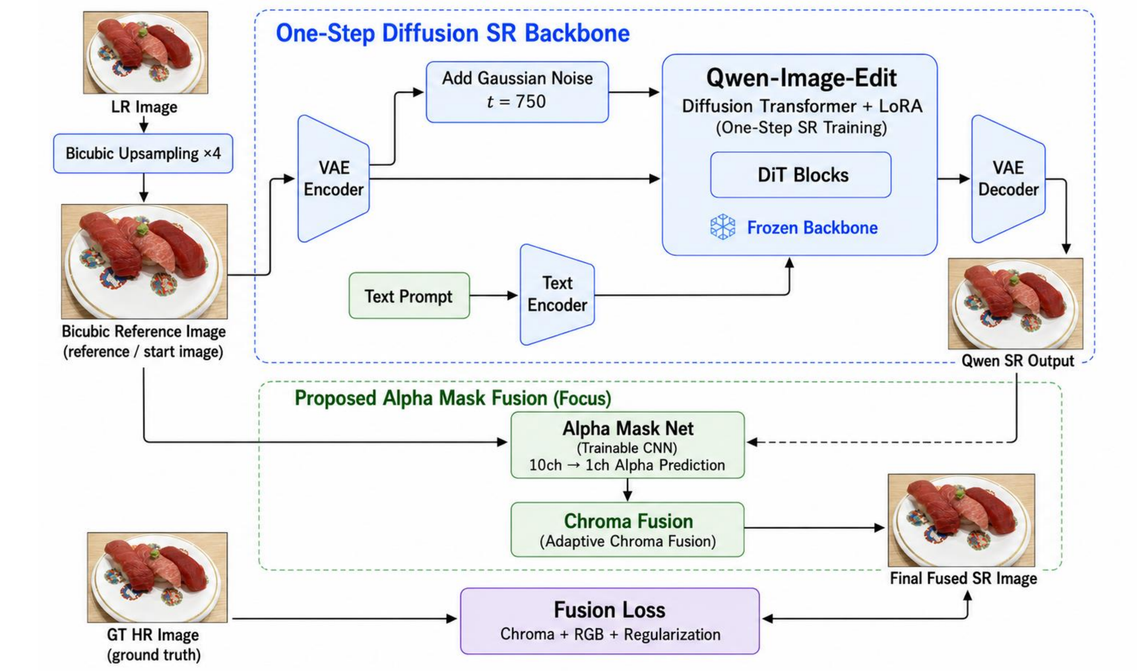
研究背景・目的

- 問題：既存の1-step拡散超解像は高速である一方、高周波テクスチャの生成時に色ずれが生じやすく、文字・標識・物体色の正確な復元が難しい。
- 方法：Qwen-Image-Editを1-step超解像モデルとして学習し、Alpha Maskで超解像画像とbicubic参照画像の色度を適応的に融合する。
- 効果：1-step推論で画質を改善し、RealSRではLPIPS 0.1949、FID 82.13、PSNR 26.31と、3指標で比較手法中最高性能を達成。色差も低減した。



提案手法

1. 全体構成



Stage 1 : One-Step Diffusion Transformer SR

入力 → bicubic 参照画像を VAE latent に変換し、固定時刻 t = 750 でノイズ化。
学習 → Qwen-Image-Edit の attention / MLP を LoRA (rank 16) で微調整し、latent L2 + LPIPS で 1-step 復元を学習。
出力 → 高周波細部を含む Qwen SR 画像。Stage 2 では backbone を凍結する。

Stage 2 : Adaptive Chrominance Fusion

入力 → Stage 1 の Qwen SR 画像 + 色の基準となる bicubic 参照画像。
処理 → 2 画像の色ずれと明るさを画素ごとに比較し、どちらの色を信頼するかを Alpha Mask として予測。
出力 → Qwen SR の輝度と細部を保ち、信頼できる色度だけを局所融合。軽量 4-layer CNN で追加計算も小さい。

2. YCbCr 色度融合

$$Y_f = Y_q$$ 輝度を保持

$$Cb_f = \alpha \odot Cb_q + (1 - \alpha) \odot Cb_r$$

$$Cr_f = \alpha \odot Cr_q + (1 - \alpha) \odot Cr_r$$

α ≈ 1 : Qwen 色度を採用　/　α ≈ 0 : 参照画像の色度を採用

3. 学習と推論

- 二段階学習：まず Stage 1 で LoRA のみを最適化して 1-step SR を獲得し、その後 backbone を凍結して Alpha Mask Net のみを学習。
- 損失設計：Stage 1はlatent L2＋LPIPS、Stage 2は色度／RGB／正則化／教師・オラクル制約。固定置換せず、色信頼度を画素単位で学習。
- 推論効率：反復拡散を用いず、1-step SR backbone と軽量 Alpha Mask Net の直列構成だけで色補正を低コストに追加できる。

実験結果

1. 定量結果

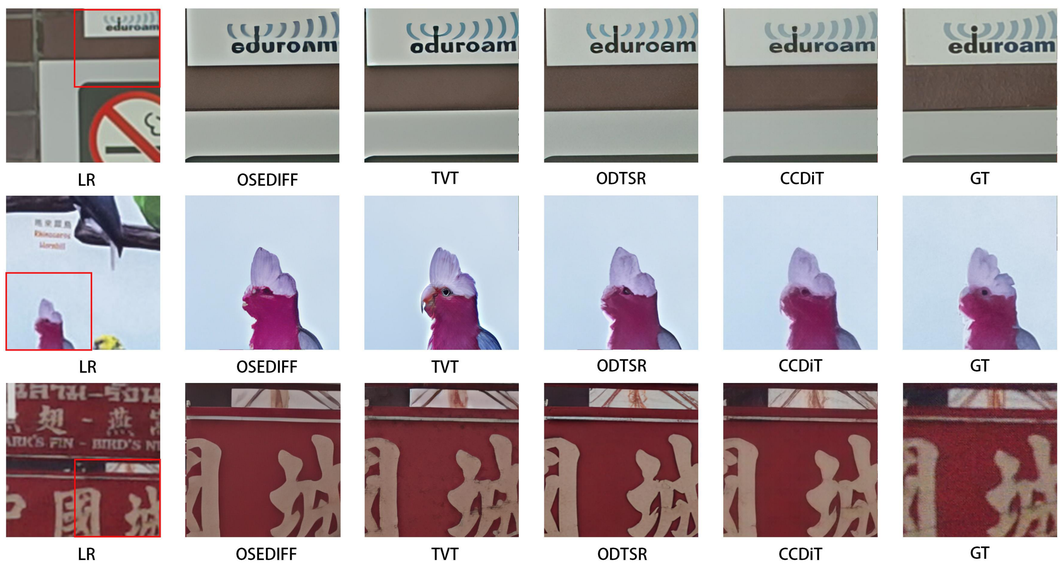
代表的な単一ステップ手法との比較

Dataset	Method	LPIPS↓	FID↓	PSNR↑	SSIM↑
RealSR	OSEDiff[2]	0.2921	123.49	**25.15**	0.7341
RealSR	ODTSR[1]	**0.2398**	**101.49**	25.07	**0.7361**
RealSR	**CCDiT**	**0.1949**	**82.13**	**26.31**	**0.7553**
DRealSR	OSEDiff[2]	0.2968	135.29	27.92	**0.7835**
DRealSR	ODTSR[1]	**0.2592**	**119.86**	**28.14**	0.7736
DRealSR	**CCDiT**	**0.2347**	**109.89**	**28.67**	**0.7766**
DIV2K-Val	OSEDiff[2]	0.2942	**26.34**	23.72	**0.6109**
DIV2K-Val	ODTSR[1]	**0.2761**	**21.47**	**23.92**	0.6108
DIV2K-Val	**CCDiT**	**0.2648**	27.14	**24.56**	**0.6178**

赤字：掲載3手法内の最良　　下線：第2位

主要結果
CCDiT は RealSR / DRealSR で LPIPS / FID / PSNR を改善。DIV2K-Val では LPIPS と PSNR が最良だが、FID など一部の知覚 / 無参照指標では既存手法が優位。

2. 定性評価



観察
既存手法は細部を回復する一方、標識や文字、高彩度境界で色や内容がGTからずれる。CCDiTは細部を保ち、GTへの忠実度を高める。

3. 色融合アブレーション

固定色度置換と適応的 Alpha Mask の比較

Model	Method	LPIPS↓	DISTS↓	FID↓	Mean ΔE↓
CCDiT	w/o Fusion	**0.1926**	**0.1777**	**79.80**	**4.6296**
CCDiT	Fixed CbCr	0.2076	**0.1758**	92.99	4.7790
CCDiT	**Alpha Mask**	**0.1949**	**0.1733**	**82.13**	**4.5566**
OSEDiff[2]	w/o Fusion	**0.2921**	0.2128	**123.50**	5.1556
OSEDiff[2]	Fixed CbCr	0.2954	**0.2106**	140.46	**5.0414**
OSEDiff[2]	**Alpha Mask**	**0.2888**	**0.2104**	**129.16**	**4.9010**

赤字：各データセット内の最良　　下線：第2位

CCDiT 内：なぜ適応融合が必要か
Fixed CbCr は参照色を一律に信頼するため、LPIPS / FID が悪化する。一方、Alpha Mask は領域ごとに色の信頼度を判断し、DISTS / Mean ΔE2000 を改善する。

OSEDiff への転用：backbone をまたぐ色補正
Alpha Maskを後段に追加するだけで、LPIPS / DISTS / Mean ΔE2000 が改善し、異なる SR backbone でも色差低減効果を確認した。

結論と限界

1-step
Qwen SR + 軽量 CNN

Color
RealSR：DISTS / ΔE 改善

Limit
DRealSR：参照色の信頼性

CCDiT は 1-step SR の再構成能力を保ちつつ、適応的な色度融合で色ずれを抑制する。Alpha Mask は 異なる SR backbone でも LPIPS / DISTS / Mean ΔE2000 を改善した。今後は参照画像の信頼度を同時推定し、劣化した参照色にも頑健な融合へ拡張する。

[1] Fang et al., ODTSR: One-Step Diffusion Transformer for Controllable Real-World Image Super-Resolution, CVPR 2026.
[2] Wu et al., OSEDiff: One-Step Effective Diffusion Network for Real-World Image Super-Resolution, NeurIPS 2024.
[3] Yi et al., TVT: Fine-structure Preserved Real-world Image Super-resolution via Transfer VAE Training, ICCV 2025.