Optimal Transport for Rectified Flow Image Editing

A Unified Framework for Inversion-Based and Inversion-Free Editing

Marian Lupascu and Mihai-Sorin Stupariu

WACV 2026 — Paper #1419

github.com/marianlupascu/OT-RF

Why is Rectified Flow Editing Hard?

Rectified Flow models (FLUX, SD3) use straight-line trajectories — great for generation, but editing is challenging:

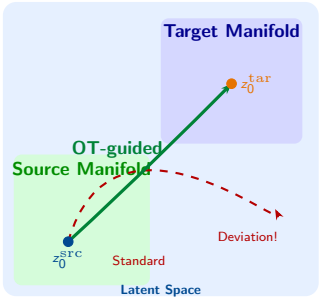
Problem 1 — Inversion-based

Deterministic ODE ⇒ no stochastic regularization.
Errors compound ⇒ **trajectory deviation**.

Problem 2 — Inversion-free (FlowEdit)

Pathways constructed *without* principled guidance.
⇒ **Suboptimal paths & semantic drift**.

Core trade-off: Reconstruction fidelity *vs.* Editing flexibility.



Our Solution: Optimal Transport Guidance

Key Insight: OT theory gives principled, minimal-cost pathways — for *both* editing paradigms, *training-free*.

Transport-Guided Inversion

OT correction at each denoising step:

$$v_{\text{enh}} = v_{\text{RF}} + \alpha(t) \cdot d_{\text{OT}}$$

Adaptive cosine schedule $\alpha(t)$. No retraining. $O(n)$ cost.

Transport-Enhanced FlowEdit

OT augmentation of FlowEdit velocity:

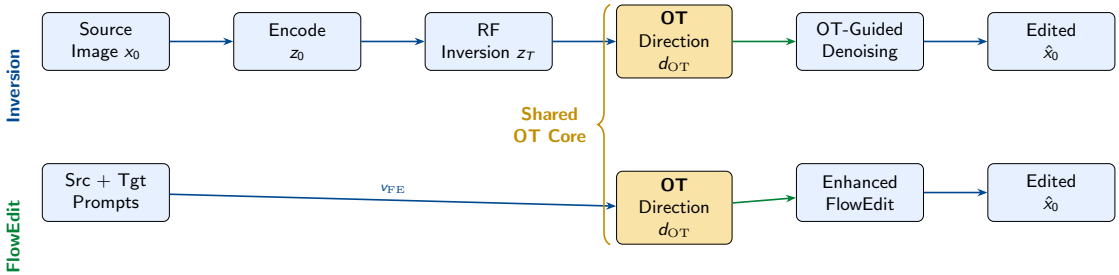
$$v_{\text{enh}} = v_{\text{FE}} + \gamma(t) \cdot d_{\text{OT}}$$

Better semantic consistency. Training-free.

Closed-Form Transport Direction (Brenier map, W_2, $O(n^3) \to O(n)$)

$$d_{\text{OT}}(z_t, z_{\text{tgt}}, t) = \frac{z_{\text{tgt}} - z_t}{\max(T - t,\ \delta)}$$

Unified OT Framework — Overview



One transport core, two paradigms. OT guidance acts as *elastic constraints* keeping edits on the source manifold.

Reconstruction Quality — SFHQ Dataset (null prompt)

Method	LPIPS↓	SSIM↑	PSNR↑	Face Rec.↓	CLIP-I↑	RT (s)↓
RF-Inversion	0.135	0.833	32.58	0.387	0.936	28.8
RF-Inv.+OTC (ours)	**0.001**	**0.992**	**40.71**	**0.112**	**0.999**	**28.6**
FlowEdit (SD3)	0.001	0.990	40.78	0.094	0.998	16.5
FlowEdit (SD3)+OTC (ours)	**0.001**	0.990	40.78	**0.093**	0.998	16.6
FlowEdit (FLUX)	0.096	0.879	24.26	0.304	0.868	28.6
FlowEdit (FLUX)+OTC (ours)	**0.001**	**0.993**	**40.65**	**0.113**	**0.999**	28.7

RF-Inversion+OTC

×135 LPIPS **+8.1 dB** PSNR **−71%** face dist.

FlowEdit (FLUX)+OTC

×96 LPIPS **+16.4 dB** PSNR **−63%** face dist.

Stroke-to-Image & Semantic Face Editing

Stroke-to-Image Reconstruction

(LSUN-Bedroom / Church, L2 ↓)

Method	Bedroom	Church
RF-Inversion	82.55	80.35
+OTC	**76.10**	**69.97**
gain	**−7.8%**	**−12.9%**
FlowEdit (FLUX)	61.51	60.76
+OTC	**57.61**	**57.43**

Semantic Face Editing

Prompt: *"... wearing glasses"*

Method	Face Rec.↓	CLIP-T↑
RF-Inversion	0.621	0.281
+OTC	**0.552**	**0.285**
gain	**−11.2%**	—
FlowEdit (FLUX)	0.537	0.293
+OTC	**0.521**	**0.296**

✓ Better identity preservation

✓ Stronger text–image alignment

Qualitative Results — Reconstruction & Stroke-to-Image

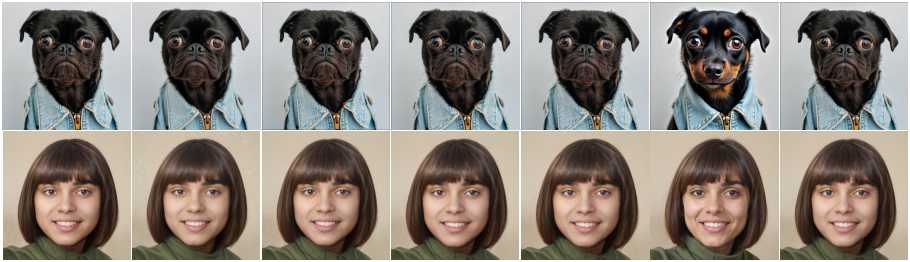


Fig. 3: Reconstruction across inversion-based (RF-Inversion, +OTC) and inversion-free (FlowEdit SD3/FLUX ±OTC) methods.

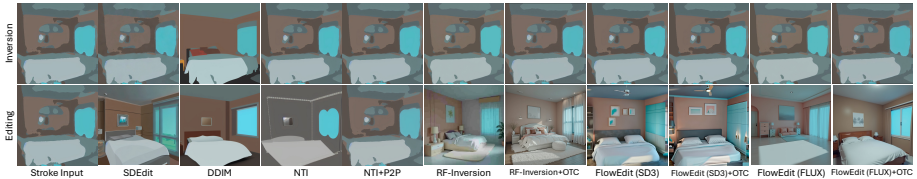


Fig. 4: Stroke-to-image: null prompt (top) vs. "a photo-realistic picture of a bedroom" (bottom).

✓ OTC consistently restores fidelity across **both inversion and FlowEdit** pipelines

Qualitative Results — Semantic Face Editing

Fig. 6: Prompt "face of a man/woman wearing glasses" — identity preserved across 3 subjects, all methods.

Fig. 7: RF-Inversion (top) vs. RF-Inversion+OTC (bottom): expression, age, gender, object insertion.

✓ Better identity preservation ✓ Expression, age, gender, objects — all improved with OTC

Key Design Choices

Closed-Form W_2 Transport

Brenier map: optimal displacement is

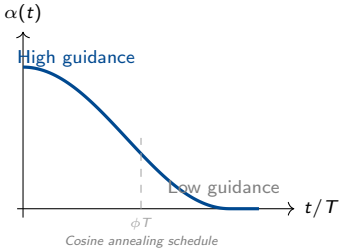
$$d_{\mathrm{OT}} = (z_{\mathrm{tgt}} - z_t) / (T - t)$$

- ✓ No iterative solver
- ✓ $O(n^3) \rightarrow O(n)$ complexity
- ✓ Numerically stable ($\delta = 0.01$)

Adaptive Cosine Schedule $\alpha(t)$

High guidance early (structure) $\rightarrow$ fades to prompt conditioning.

- ✓ Smooth, differentiable
- ✓ Phase param ϕ controls scope



Gradient Clipping

$\|d_{\mathrm{OT}}\|_2 \leq \tau = 10.0$

Prevents OT from overwhelming RF velocity.

Summary

Contributions

(i) OT distances $\leftrightarrow$ RF editing quality

(ii) **Transport-Guided Inversion** with cosine scheduling

(iii) **Transport-Enhanced FlowEdit**

(iv) Validated on **FLUX** & **SD3**

(v) **Training-free** — zero computational overhead

Key Numbers

- $\times$**135** LPIPS — RF-Inversion+OTC
- $+$**16.4 dB** PSNR — FlowEdit (FLUX)+OTC
- $-$**11.2%** face distance — semantic editing
- $-$**12.9%** L2 — LSUN-Church stroke

Take-Away

Optimal Transport provides a **unified, training-free,** $O(n)$ framework for rectified flow image editing across *both paradigms*.

Code & Results

`github.com/marianlupascu/OT-RF`

Thank you! Questions?

Algorithm 1: Transport-Guided RF Inversion (simplified)

Require: Source x_0, prompt p, model v_θ, β_0, ϕ

$z_0 \leftarrow \mathrm{Encode}(x_0)$; $z_{\mathrm{tgt}} \leftarrow z_0$; $z_T \leftarrow \mathrm{RFInvert}(z_0, v_\theta)$

for $t = T$ **down to** 0, step $-\Delta t$ **do**

  $v_{\mathrm{tar}} \leftarrow v_\theta(z_t, t, p)$; $v_{\mathrm{ref}} \leftarrow v_\theta(z_t, t, \varnothing, z_0)$

  $v_{\mathrm{RF}} \leftarrow v_{\mathrm{tar}} + \eta(v_{\mathrm{ref}} - v_{\mathrm{tar}})$

  $d_{\mathrm{OT}} \leftarrow (z_{\mathrm{tgt}} - z_t)/\max(T - t, \delta)$

  $\alpha \leftarrow \beta_0 \cdot \frac{1}{2}(1 + \cos(\min(\frac{T-t}{T\phi}, 1)\pi))$

  $v_{\mathrm{enh}} \leftarrow v_{\mathrm{RF}} + \alpha \cdot \mathrm{clip}(d_{\mathrm{OT}}, \tau)$

  $z_{t-\Delta t} \leftarrow z_t + \Delta t \cdot v_{\mathrm{enh}}$

end for

return $\mathrm{Decode}(z_0)$