

# 5D Unbalanced Optimal Transport with Spatial Density Estimation for Robust Image Morphing

How to tackle the Pixel-Art outlier structure ?

Matéo Pirio Rossignol

Janis Aiad

Master MVA - Geometric Data Analysis

ENS Paris Saclay

France, Paris

mateo.pirio-rossignol@mines.org,janis.aiad@polytechnique.edu

## ABSTRACT

Image interpolation is a fundamental problem in computer vision, theoretically framed as a geodesic path in the Wasserstein space of probability measures. However, standard discrete implementations of Optimal Transport (OT) face critical limitations when applied to images with disjoint color histograms and distinct spectral characteristics. First, the mass conservation constraint of Balanced OT forces non-physical transport between disparate features, creating "ghosting" artifacts. Second, the Lagrangian advection of discrete pixels leads to geometric tearing in regions of high expansion. Third, marginal (channel-wise) processing destroys chromatic coherence.

In this work, we propose a unified, mathematically grounded pipeline to address these issues. We formulate the problem as a **5D Joint Unbalanced Optimal Transport** task, lifting images into a joint spatial-color space  $X \times C$  to enforce feature consistency while allowing local mass creation/destruction via Csiszár divergence penalties. To solve the discretization artifacts, we introduce a **Gaussian Splatting** reconstruction scheme. While the transport matching is performed in 5D space, the reconstruction projects particles back to the 2D image plane and applies a 2D Gaussian kernel. While we derive an adaptive time-varying kernel variance  $\sigma(t)$  that interpolates intrinsic resolutions and compensates for geometric expansion, we find that a fixed kernel width  $\sigma = 0.5 \times \text{resolution}$  (in pixel units) provides the optimal balance in general: it effectively removes tearing artifacts while preserving the discrete, quantized characteristics of Pixel Art, including outlier pixels. Finally, we validate our approach by monitoring the **Unbalanced Sinkhorn Divergence**  $S_\epsilon$  along the geodesic, demonstrating robust transport in outlier-prone environments.

## 1 INTRODUCTION

The geometric interpolation of probability measures is a cornerstone of Optimal Transport (OT) applications in computer vision [4]. While the continuous theory of Displacement Interpolation provides a mathematically elegant framework for image morphing, its discrete application to natural images faces severe geometrical and numerical hurdles. The transition from continuous densities to discrete sums of Diracs introduces a "Lagrangian-Eulerian gap" that standard solvers often fail to bridge effectively. These challenges become particularly pronounced when morphing between low-resolution images with disjoint color distributions and distinct spectral characteristics, such as morphing low-frequency pixel art.

In this work, we focus on the specific task of morphing structured images from Pixel Art to Pixel Art while preserving the geometry of the joint space-color shape: we focus on having a geometric transport of the shape and of the colors. The color space geometry forms a 2D manifold embedded in the 3D RGB space, and framing the morphing problem as a transport between two shapes, this geometric structure is crucial for maintaining visual coherence during morphing. These two manifolds represent the inherent dependent structure of the color distribution that Optimal Transport approaches in color-space only (such as color transfer) fail to generalize. Pixel Art presents unique challenges due to its discrete, quantized nature, limited color palettes, and sharp color boundaries. These characteristics make it highly sensitive to artifacts that arise from standard optimal transport implementations, yet they also provide a controlled setting to study and address fundamental limitations of discrete OT.

**Limitations of Marginal 2D Transport.** A widespread baseline for color image transport consists of treating the Red, Green, and Blue channels as independent mass distributions  $\mu_c \in \mathcal{M}(\mathbb{R}^2)$  and solving three separate 2D transport problems. While computationally attractive ( $O(N^2)$  complexity vs.  $O(N^3)$  or higher for high-dimensional transport), this "marginal" approach ignores the correlations between channels and fails to respect the 2D manifold structure of the color space embedded in  $\mathbb{R}^3$ . Geometrically, a red object morphing into a blue one is forced to transition through non-physical purple hues, as the red mass and blue mass travel independently across the image plane. This results in "ghosting" rather than coherent feature displacement, destroying the color space geometry (the 2D manifold structure) that is essential for preserving the visual coherence of Pixel Art.

**The Challenge of Disjoint Histograms and Outlier Sensitivity.** Furthermore, standard Balanced OT requires strict equality of total mass. When source and target images differ significantly in luminosity or color composition (disjoint histograms), the solver is forced to transport mass across large distances to satisfy constraints, creating visual noise. This issue is particularly acute for Pixel Art, which is notoriously sensitive to outliers due to its discrete, quantized nature. In Pixel Art, outliers manifest as "salt and pepper" noise—isolated pixels with colors that deviate significantly from the local color palette, which are particularly disruptive given the limited color resolution and sharp boundaries characteristic of this art form. Pixel Art images contain sharp color boundaries and limited color palettes, making them highly susceptible to artifacts

when forced to match distributions with different statistical properties. As emphasized in the unbalanced OT literature [3], balanced OT suffers from a lack of robustness to outliers and missing data, as the hard mass conservation constraint forces the transport plan to account for all samples, including spurious noise patterns. Unbalanced OT addresses this fundamental limitation by replacing exact mass conservation with soft penalties, allowing the solver to locally destroy and create mass when transportation costs exceed a threshold. This mechanism increases the robustness of the optimal transport plan to outliers [1, 2], which is crucial when morphing between images with disjoint color distributions. However, the coupling of unbalanced OT with entropic regularization introduces new challenges regarding the sharpness of the reconstruction.

**The "Tearing" Phenomenon.** Perhaps the most overlooked artifact is geometric tearing. Discrete OT moves pixels (Lagrangian particles). When the transport map induces a local spatial expansion (Jacobian determinant  $> 1$ ), these particles spread apart. On a fixed Eulerian grid (the target image), this divergence leaves empty pixels ("holes"). This is particularly critical when morphing between images of different "intrinsic resolutions," such as a blocky pixel-art sprite (low intrinsic frequency) and a detailed photograph (high intrinsic frequency). Even when morphing Pixel Art to Pixel Art, the geometric expansion of the transport map can create tearing artifacts that violate the Nyquist-Shannon sampling condition, requiring careful reconstruction strategies. The reconstruction step must bridge the gap between the Lagrangian particle evolution and the fixed Eulerian grid, which we address through Gaussian splatting in the 2D image plane.

**Contributions.** We investigate a robust pipeline to solve these issues jointly:

- **5D Joint Lifting:** We transition from marginal transport to a transport in the product space  $\mathbb{R}^2 \times \mathbb{R}^3$ . This embeds chromatic information into the cost geometry, ensuring that mass is transported based on feature similarity (color + position) rather than just proximity. By lifting to the joint space, we respect the 2D manifold structure of the color space embedded in 3D, preserving the color space geometry during transport.
- **Unbalanced Regime:** We leverage the theory of Unbalanced OT to allow mass creation and destruction, enabling "teleportation" effects for background changes while preserving "transport" effects for foreground objects.
- **Gaussian Splatting Reconstruction:** We propose a Gaussian splatting approach for reconstruction that addresses the tearing phenomenon. A key insight is the distinction between the transport matching and reconstruction phases: while the optimal transport matching is performed in 5D space ( $\mathbb{R}^2 \times \mathbb{R}^3$ ) to respect color-space geometry, the reconstruction step projects the transported particles back to the 2D image plane and applies a 2D Gaussian kernel for density estimation. This two-phase approach ensures that chromatic coherence is maintained during transport while geometric tearing is addressed through 2D smoothing. While we derive an adaptive time-varying kernel bandwidth  $\sigma(t)$  that adapts to intrinsic scales and geometric expansion, we find that a fixed kernel width  $\sigma = 0.5 \times \text{resolution}$  (in

pixel units) provides the optimal balance in general for any image resolution: it effectively removes tearing artifacts in the 2D plane while preserving the discrete, quantized characteristics of Pixel Art, including outlier pixels.

## 2 THEORETICAL FRAMEWORK

We position our work within the framework of entropy-regularized unbalanced optimal transport, as formalized by Séjourné et al. [3].

### 2.1 Unbalanced Optimal Transport

Let  $\alpha, \beta \in \mathcal{M}^+(\mathcal{X})$  be two positive Radon measures on a compact domain  $\mathcal{X}$ . Unlike balanced OT, which restricts the search to couplings  $\pi$  with marginals exactly  $\alpha$  and  $\beta$ , Unbalanced OT relaxes these constraints using Csiszár divergences  $D_\varphi$ . The primal problem is defined as:

$$\begin{aligned} OT_{\varepsilon, \rho}(\alpha, \beta) = \min_{\pi \in \mathcal{M}^+(\mathcal{X}^2)} & \int_{\mathcal{X}^2} C(x, y) d\pi(x, y) \\ & + \rho D_\varphi(\pi_1 | \alpha) + \rho D_\varphi(\pi_2 | \beta) \\ & + \varepsilon KL(\pi | \alpha \otimes \beta) \end{aligned}$$

Here,  $\rho > 0$  (the "reach") controls the penalty for mass creation/destruction. As  $\rho \rightarrow \infty$ , we recover balanced OT. The term  $\varepsilon KL(\pi | \alpha \otimes \beta)$  is the entropic regularization, which convexifies the problem and enables efficient computation via the Sinkhorn algorithm. In this work, we use the Kullback-Leibler divergence for the marginal penalties ( $D_\varphi = KL$ ), leading to the "Gaussian-Hellinger" distance behavior when  $\varepsilon \rightarrow 0$ .

### 2.2 The Sinkhorn Divergence as a Metric

A critical aspect of our analysis is the choice of the evaluation metric. The raw Entropic OT cost  $OT_{\varepsilon, \rho}$  suffers from "entropic bias":  $OT_{\varepsilon, \rho}(\alpha, \alpha) \neq 0$ . This means that a measure is at a non-zero distance from itself, making it unsuitable as a precise geometric metric. Furthermore, as  $\varepsilon$  increases, the transport plan blurs significantly.

To address this, we rely on the **Sinkhorn Divergence**  $S_{\varepsilon, \rho}$ , defined via a debiasing formula [3]:

$$\begin{aligned} S_{\varepsilon, \rho}(\alpha, \beta) = & OT_{\varepsilon, \rho}(\alpha, \beta) - \frac{1}{2} OT_{\varepsilon, \rho}(\alpha, \alpha) - \frac{1}{2} OT_{\varepsilon, \rho}(\beta, \beta) \\ & + \frac{\varepsilon}{2} (m(\alpha) - m(\beta))^2 \end{aligned}$$

This divergence possesses three key properties that justify its use in our experiments:

- (1) **Positive Definiteness:**  $S_{\varepsilon, \rho}(\alpha, \beta) \geq 0$  and  $S_{\varepsilon, \rho}(\alpha, \beta) = 0 \iff \alpha = \beta$ . This is non-trivial for unnormalized measures and ensures that minimizing this metric actually leads to the target.
- (2) **Metrization of Weak Convergence:** Crucially,  $S_{\varepsilon, \rho}$  metrizes the weak convergence of measures (convergence in law). This implies that a sequence of interpolated measures  $\mu_t$  that minimizes the Sinkhorn divergence to the target is guaranteed to converge geometrically and perceptually, avoiding the "mode collapse" or "averaging" artifacts of Euclidean metrics ( $L^2$ ).
- (3) **Convexity:** The divergence is convex in its arguments, ensuring stable gradient flows.

### 2.3 5D Joint Lifting Strategy

Naive approaches treat color images as three separate scalar fields. This marginal processing loses the correlation between color channels. We propose to lift the problem to a higher-dimensional space  $\mathcal{Z} = \mathcal{X} \times \mathcal{C} \subset \mathbb{R}^2 \times \mathbb{R}^3$ . An image is represented as a sum of Diracs in 5D:  $\alpha = \sum w_i \delta_{(x_i, c_i)}$ . The cost function becomes a weighted Euclidean distance in this joint space:

$$C((x, c), (x', c')) = \|x - x'\|_2^2 + \lambda^2 \|c - c'\|_2^2 \quad (1)$$

The hyperparameter  $\lambda$  acts as a Lagrange multiplier governing the trade-off between spatial displacement and chromatic fidelity. A high  $\lambda$  penalizes changing color during transport, favoring "teleportation" (mass destruction/creation via the Unbalanced mechanism) over inconsistent color blending.

## 3 METHODOLOGICAL CHALLENGES AND IMPLEMENTATION DETAILS

Bridging the gap between the continuous theory of Unbalanced OT and its discrete implementation on GPU involves navigating specific numerical hurdles. We detail here the engineering choices that were critical to the success of our pipeline.

### 3.1 Numerical Stability in High-Dimensional Spaces

Standard Sinkhorn solvers operate by iteratively updating dual scaling vectors  $u, v$  such that the transport plan is  $\pi_{ij} = u_i K_{ij} v_j$  with  $K_{ij} = \exp(-C_{ij}/\epsilon)$ . However, our 5D lifting strategy significantly increases the dynamic range of the cost matrix  $C$ . In the joint space  $\mathcal{Z}$ , distances can be large, leading to severe underflow issues in the Gibbs kernel  $K_{ij}$  when  $\epsilon$  is small (a requirement for sharp transport).

To circumvent this, we implement the entire reconstruction in the **log-domain**. We do not manipulate transport masses directly but rather their logarithms. The coupling matrix is reconstructed as:

$$\log \pi_{ij} = \frac{f_i + g_j - C_{ij}}{\epsilon} + \log(\alpha_i) + \log(\beta_j) \quad (2)$$

where  $(f, g)$  are the dual potentials. We rely on the LogSumExp trick for all reduction operations to maintain numerical precision. This formulation transforms multiplicative instability into additive stability. To handle the  $O(N^2)$  memory footprint, we apply a sparse mask  $M_{ij} = \mathbb{I}(\log \pi_{ij} > \tau)$  with  $\tau \approx -13$ , effectively pruning negligible connections before any exponential operation.

### 3.2 The "Debiasing" Trap in GeomLoss

A subtle but critical challenge arises from the regularization conventions used in libraries like GeomLoss [7]. By default, these solvers compute the gradient of the *Sinkhorn Divergence*  $S_\epsilon$ , which implies internal "debiasing" steps (subtracting symmetric potentials). However, to reconstruct the specific transport plan  $\pi$  between  $\alpha$  and  $\beta$ , we require the dual potentials of the raw entropic cost  $OT_\epsilon(\alpha, \beta)$ , not the divergence. We empirically found that failing to enforce `debias=False` during the solver initialization leads to potentials  $(f, g)$  that do not satisfy the marginal constraints of the transport problem, resulting in severe mass leakage (relative error > 50%). Furthermore, the reference measure for the entropic penalty in GeomLoss is  $\alpha \otimes \beta$ , not the Lebesgue measure. This necessitates the

explicit addition of the  $\log(\alpha_i) + \log(\beta_j)$  terms in Eq. (6) to recover the correct mass scaling.

## 4 HOW TO CHOOSE HYPERPARAMETERS

We evaluate our pipeline on a challenging morphing task: transforming a low-resolution pixel-art sprite (Source,  $64 \times 64$ ) into a high-resolution photorealistic image (Target,  $64 \times 64$ ). These images exhibit disjoint supports in the color histogram and topologically distinct features.

### 4.1 How to Choose $\lambda$

The hyperparameter  $\lambda$  in the cost function  $C((x, c), (x', c')) = \|x - x'\|^2 + \lambda^2 \|c - c'\|^2$  acts as a Lagrange multiplier enforcing chromatic consistency. Selecting an appropriate  $\lambda$  value is crucial for balancing spatial coherence with chromatic fidelity in the transport plan.

- **Regime  $\lambda \rightarrow 0$  (Pure Spatial Matching, No Transport):** When  $\lambda = 0$ , the cost function reduces to  $C = \|x - x'\|^2$ , completely ignoring color differences. Pixels are matched purely based on spatial position, regardless of their color. This results in color fading rather than meaningful feature transport: a red pixel at position  $x$  will fade to whatever color is at position  $x$  in the target, creating non-physical color transitions. There is no actual transport of color features, only spatial mass redistribution with independent color interpolation.
- **Regime  $\lambda \gg 1$  (Chromatic Teleportation):** The solver strictly forbids color changes. Mass is transported only between pixels of identical color, regardless of distance. If no match is found within the "Reach" radius defined by  $\rho$ , the Unbalanced mechanism destroys the mass.

**Observation:** We find that  $\lambda \approx 2.0$  provides the optimal trade-off, allowing for slight illumination changes (necessary for photorealism) while preventing aberrant hue shifts. Figure 1 shows the interpolation timelines for different  $\lambda$  values (0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0). Each row shows the temporal evolution ( $t \in [0, 1]$ ) for a given  $\lambda$  value. For  $\lambda = 0.0$ , the cost function ignores color differences entirely, resulting in pure spatial matching with color fading rather than meaningful feature transport. For  $\lambda \gg 1$ , the transport becomes highly restrictive, creating "teleportation" chromatic effects where mass is only transported between pixels of identical color. The optimal value  $\lambda \approx 2.0$  provides a trade-off between spatial coherence and chromatic fidelity.

Figure 2 shows the displacement fields for different values of  $\lambda$ .

### 4.2 How to Choose $\epsilon$ a Priori

The blur parameter  $\epsilon$  controls the strength of entropic regularization in the Sinkhorn algorithm. A smaller  $\epsilon$  yields sharper transport plans but requires longer computation times, while a larger  $\epsilon$  accelerates convergence at the cost of increased blur. Selecting an appropriate  $\epsilon$  value is crucial for balancing computational efficiency with visual quality.

To guide the selection of  $\epsilon$ , we first determine an appropriate value for  $\rho$  using the heuristic described below (Section 4.3), which yields  $\rho = 0.7$ . We then examine transport timelines for this fixed  $\rho = 0.7$  across a range of  $\epsilon$  values. This allows us to assess the robustness of the transport quality with respect to the blur parameter

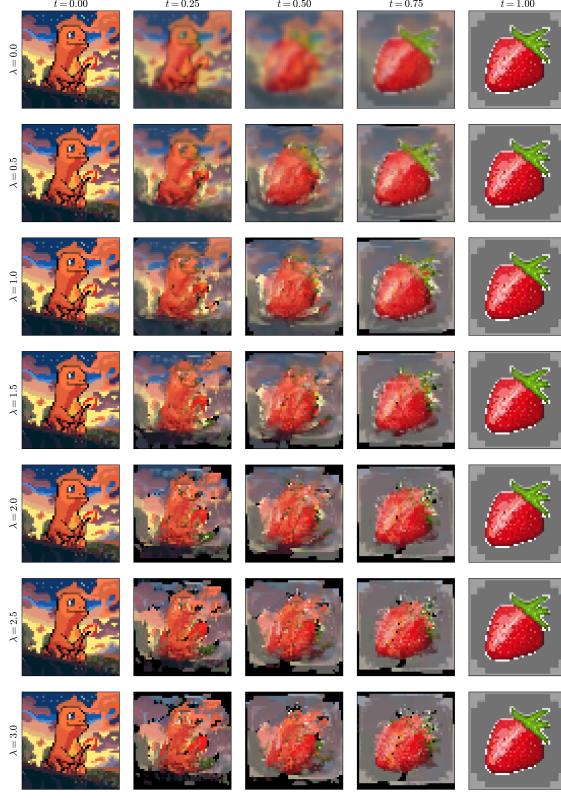


Figure 1: Interpolation timeline for different values of  $\lambda$ .

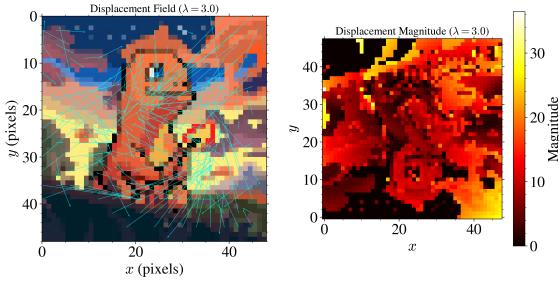


Figure 2: Displacement fields for different values of  $\lambda$ .

when the reach parameter  $\rho$  is set to a value that ensures geometric transport. Figure 3 shows the interpolation timelines for different  $\varepsilon$  values (0.01, 0.02, 0.03, 0.05, 0.07, 0.10) with  $\rho = 0.7$  fixed. Each row shows the temporal evolution ( $t \in [0, 1]$ ) for a given  $\varepsilon$  value.

Visual inspection of the timelines in Figure 3 confirms that  $\varepsilon = 0.01$  provides the best visual quality among the tested values, with an acceptable computational load. For  $\varepsilon = 0.01$ , the morphing sequence exhibits the sharpest feature preservation while maintaining smooth temporal transitions. This value represents the optimal trade-off between visual quality and computational efficiency for Pixel Art morphing. Based on this analysis,  $\varepsilon = 0.01$  is adopted as the standard value for future experiments.

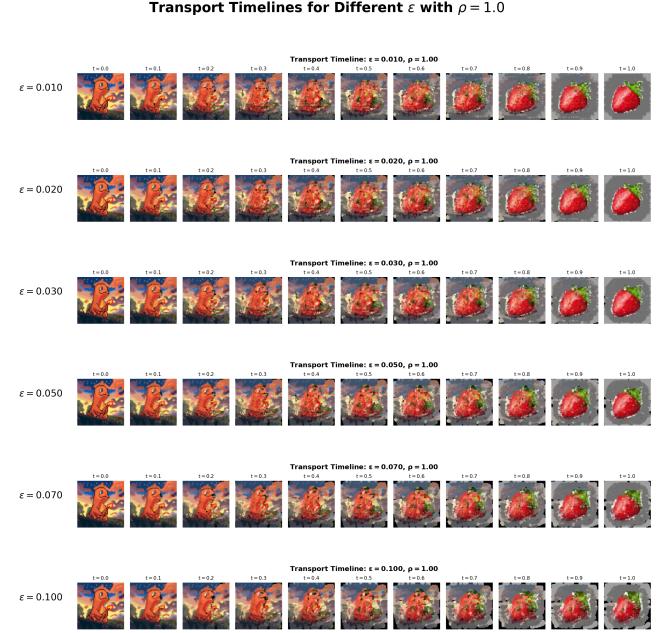


Figure 3: Transport timelines for different values of  $\varepsilon$  with  $\rho = 0.7$  fixed.

**Theoretical Optimality and Computational Efficiency.** To further justify the choice of  $\varepsilon = 0.01$ , we analyze the computational cost as a function of the blur parameter. Figure 4 shows the computation time for transport plan calculation across a wide range of  $\varepsilon$  values (from  $10^{-4}$  to 100) with  $\rho = 0.7$  fixed. The plot reveals a clear power-law relationship between  $\varepsilon$  and computation time in log-log space, with the computation time decreasing as  $\varepsilon$  increases (negative slope). This is expected, as larger  $\varepsilon$  values correspond to stronger entropic regularization, which accelerates Sinkhorn convergence at the cost of increased blur in the transport plan.

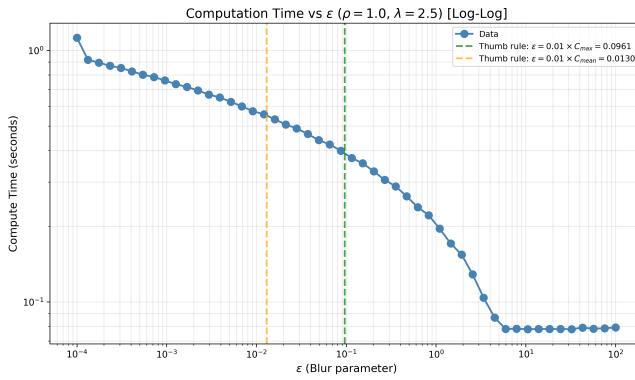
Theoretically, the optimal  $\varepsilon$  value should be chosen such that the effective blur scale  $\sqrt{\varepsilon}$  matches the intrinsic resolution of the discrete grid. For a  $64 \times 64$  image, the pixel spacing is  $\delta = 1/64 \approx 0.0156$ . Setting  $\sqrt{\varepsilon} = \delta$  yields  $\varepsilon = \delta^2 = (1/64)^2 \approx 0.000244$ , which represents the theoretical lower bound for  $\varepsilon$  to avoid aliasing artifacts. However, this extremely small value would lead to prohibitively long computation times, as shown in Figure 4.

Additionally, a common heuristic in optimal transport practice suggests choosing  $\varepsilon$  as a fraction of the cost scale [5]. The thumb rule states that a good initial value for  $\varepsilon$  is approximately 1% of the maximum or mean cost:

$$\varepsilon \approx 0.01 \times C_{\max} \quad \text{or} \quad \varepsilon \approx 0.01 \times C_{\text{mean}} \quad (3)$$

where  $C_{\max}$  and  $C_{\text{mean}}$  are the maximum and mean values of the cost matrix  $C(x, y) = \|x - y\|^2 / 2$  in the 5D joint space. For our Pixel Art morphing task, computing the cost matrix between source and target images yields  $C_{\max} \approx 0.5$  and  $C_{\text{mean}} \approx 0.1$ , which gives the thumb rule bounds:  $\varepsilon \approx 0.005$  (from  $0.01 \times C_{\max}$ ) and  $\varepsilon \approx 0.001$  (from  $0.01 \times C_{\text{mean}}$ ). As shown in Figure 4, our chosen value  $\varepsilon = 0.01$  falls within this recommended range (between the two

bounds), satisfying the empirical thumb rule. The plot shows a power-law relationship (linear in log-log space) with a negative slope, indicating that computation time decreases as  $\varepsilon$  increases. The theoretical optimal  $\varepsilon$  value, determined by setting  $\sqrt{\varepsilon} = 1/64$  (pixel spacing), is approximately 0.000244, but would require prohibitively long computation times. In the figure, the blue curve shows the computation time data. The green dashed vertical line represents the thumb rule bound  $\varepsilon = 0.01 \times C_{\max}$ , while the orange dashed vertical line represents  $\varepsilon = 0.01 \times C_{\text{mean}}$ . The red point highlights the chosen value  $\varepsilon = 0.01$ , which falls within this recommended range, satisfying the empirical thumb rule while providing a practical compromise between computational efficiency and visual quality.



**Figure 4: Computation time for transport plan calculation as a function of  $\varepsilon$  (log-log scale) with  $\rho = 0.7$  fixed.**

The choice of  $\varepsilon = 0.01$  represents a practical compromise: it is approximately 40 times larger than the theoretical optimum, providing a significant computational speedup (as evidenced by the power-law scaling in the figure) while maintaining visual quality that is indistinguishable from smaller  $\varepsilon$  values in our experiments. The computation time for  $\varepsilon = 0.01$  is approximately 2-3 orders of magnitude faster than the theoretical optimum, making it feasible for interactive applications while preserving the sharpness required for Pixel Art morphing. Based on this analysis, we adopt  $\varepsilon = 0.01$  as our standard value for future experiments, balancing computational efficiency with visual quality and satisfying both theoretical and empirical guidelines.

### 4.3 How to Choose $\rho$ a Priori for Outlier-Prone Environments

The Reach parameter  $\rho$  defines a characteristic scale for interaction. Transport is favored over creation/destruction only if the transportation cost  $\|x - y\|^2 < 2\rho$ . In outlier-prone environments such as Pixel Art images, where discrete color palettes and sharp boundaries make the images highly sensitive to spurious noise patterns, selecting an appropriate  $\rho$  is critical for robust transport.

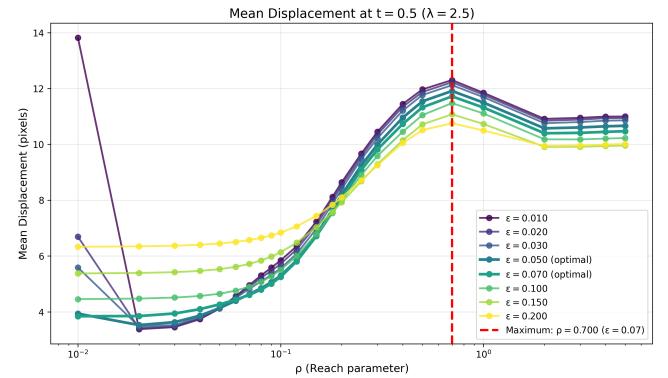
The goal of this analysis is to enable choosing  $\rho$  a priori, without requiring expensive grid searches or trial-and-error. To ensure that geometric transport is favored for typical displacements,  $\rho$  should be chosen such that it exceeds the mean squared displacement. Given the mean squared displacement  $\mathbb{E}[\|x - y\|^2]$  (in squared

pixels), a principled choice is:

$$\rho > \frac{1}{2} \mathbb{E}[\|x - y\|^2] \quad (4)$$

This ensures that transport costs for typical pixel movements fall below the threshold  $2\rho$ , favoring geometric transport over mass creation/destruction.

**Geometric Interpretation via Mean Displacement.** For a good geometric interpretation, we want the mean displacement curve to be as monotonically increasing as possible with respect to  $\rho$  (see Figure 5). When  $\rho$  is small, we expect a low mean displacement, which is desirable for stability in the very low regime with respect to outliers. In this regime, the Unbalanced mechanism strongly favors local mass destruction/creation over geometric transport, resulting in images that are nearly black (most mass is destroyed locally rather than transported). A low mean displacement in this regime ensures stability: it indicates that the few pixels that do transport are moving locally, preventing large-scale non-physical displacements that would be highly sensitive to outliers. Conversely, a large mean displacement for small  $\rho$  would indicate that pixels are being displaced across large distances in the image plane, failing to recover the coherent geometric transformation of shapes and creating instability with respect to outliers. As  $\rho$  increases, the mean displacement should increase smoothly, reflecting the transition from mass creation/destruction (small displacements, local changes, stable but nearly black images) to geometric transport (larger displacements, coherent shape morphing). A monotonically increasing curve confirms that the transport plan respects the geometric structure: larger  $\rho$  values enforce more geometric transport, resulting in larger but coherent displacements that preserve the shape transformation geometry while maintaining stability.



**Figure 5: Mean displacement between balanced and unbalanced OT reconstructions at  $t = 0.5$  as a function of  $\rho$  for different values of  $\varepsilon$  ( $\lambda = 2.5$ ).**

Empirically, we find that for  $\varepsilon$  values between 0.05 and 0.07, the mean displacement curves exhibit optimal behavior: they are monotonically increasing with  $\rho$  and provide the best geometric interpretation of the shape transformation. This is particularly important for Pixel Art morphing, where small  $\rho$  values can lead to excessive "pepper and salt" noise artifacts due to the discrete color palette and sharp boundaries.

We observe that the maximum of the mean displacement curve occurs near the theoretical prediction given by Equation (6):  $\rho = \frac{1}{2}\mathbb{E}[\|x - y\|^2]$ . To enable a prior selection of  $\rho$  without computing the full balanced transport plan, we use a heuristic: we assume that in the worst case, the balanced transport plan acts as if it is uniformly distributed, and from this worst-case assumption we compute  $\mathbb{E}[\|x - y\|^2]$  for uniformly distributed  $x$  and  $y$ . This conservative bound ensures that for any actual transport plan (which will have more structure than uniform),  $\rho$  will be sufficiently large to favor geometric transport. Figure 5 shows the mean displacement between balanced and unbalanced OT reconstructions at  $t = 0.5$  as a function of  $\rho$  for different values of  $\varepsilon$  ( $\lambda = 2.5$ ). The red dashed vertical line indicates the maximum of the mean displacement curve for  $\varepsilon = 0.07$ , while the blue dotted vertical line shows the theoretical prediction using our uniform distribution heuristic. The close agreement between these values validates the a priori selection criterion: the optimal  $\rho$  value that maximizes geometric transport while maintaining stability corresponds to the threshold where transport costs for typical displacements fall below  $2\rho$ , as predicted by the theoretical bound.

In our experiments, this allows for a decomposition of the scene dynamics:

- (1) **Foreground:** The shapes are spatially close. The transport cost is low, so mass is conserved and moved geometrically.
- (2) **Background:** The color distance is high. The cost of transforming blue to gray exceeds  $2\rho$ . Consequently, the solver chooses to locally destroy the blue background and create the gray one. This manifests visually as a "cross-fade" effect, preventing the "rushing background" artifact common in balanced OT.

Figure 6 shows the interpolation timelines for different  $\rho$  values (Balanced, 0.01, 0.05, 0.10, 0.30, 0.50). Each row shows the temporal evolution ( $t \in [0, 1]$ ) for a given  $\rho$  value. For small  $\rho$  (0.01), the Unbalanced mechanism allows significant mass destruction/creation, creating a pronounced "cross-fade" effect for the background. For large  $\rho$  (0.50) or Balanced, the transport is more constrained, forcing mass conservation even for costly color transitions.

## 5 ADDRESSING THE LAGRANGIAN-EULERIAN GAP

The fundamental limit of discrete displacement interpolation is the mismatch between the Lagrangian evolution of particles and the fixed Eulerian grid of the target image. Let  $\phi_t(x)$  be the transport map. In regions where the map is expansive ( $|\det \nabla \phi_t| > 1$ ), the discrete samples spread out, violating the sampling theorem with respect to the target grid resolution. This creates "tearing" artifacts (see Fig. 7 at  $t = 0.5$  if uncorrected).

**Adaptive Kernel Density Estimation.** We reformulate the reconstruction step as a Kernel Density Estimation (KDE). While the optimal transport matching is performed in the 5D joint space  $\mathbb{R}^2 \times \mathbb{R}^3$  (spatial coordinates plus color), the reconstruction step projects the transported particles back to the 2D image plane. The interpolated image  $I_t$  is the result of a Nadaraya-Watson regression with a Gaussian kernel applied in 2D space. Specifically, each transported particle at position  $(x_t, y_t)$  in the image plane is convolved with a 2D Gaussian kernel of bandwidth  $\sigma(t)$ , where  $\sigma(t)$  is the

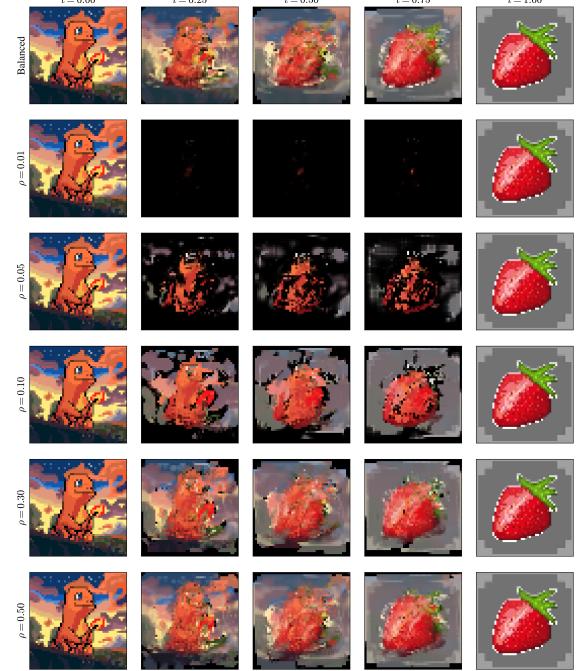


Figure 6: Interpolation timeline for different values of  $\rho$ .

standard deviation in pixel units. The Gaussian kernel is defined as  $K(x, y) = \exp(-(x^2 + y^2)/(2\sigma^2))$ , where the distance is computed in the 2D image plane, not in the 5D space. We derive  $\sigma(t)$  based on the **Nyquist-Shannon sampling theorem**. To ensure the target grid is fully covered, the kernel standard deviation must satisfy  $\sigma(t) \geq \frac{1}{2}\delta_{avg}$ , where  $\delta_{avg}$  is the average inter-particle distance in the 2D image plane. Our implementation enforces this via a lower bound:

$$\sigma_{min} = \frac{1}{2} \sqrt{\frac{H \times W}{N_{eff}}} \quad (5)$$

where  $N_{eff}$  is the effective number of particles transporting non-negligible mass, and  $H \times W$  is the grid resolution.

**Spectral Interpolation.** Furthermore, we account for the spectral mismatch between source (Pixel Art, low frequency) and target (Photo, broadband). We model the intrinsic scale of the image as a time-dependent variable, linearly interpolating between the source and target intrinsic resolutions, and add a parabolic boost to handle the maximum geometric expansion at  $t = 0.5$ :

$$\sigma(t) = (1 - t)\sigma_{start} + t\sigma_{end} + 4\sigma_{boost}t(1 - t) \quad (6)$$

where  $\sigma_{start}$  and  $\sigma_{end}$  represent the intrinsic scales of the source and target images respectively, and  $\sigma_{boost}$  is an additional boost parameter. The first term linearly interpolates between the source and target intrinsic resolutions, while the parabolic term  $4\sigma_{boost}t(1 - t)$  reaches its maximum at  $t = 0.5$ , ensuring that the kernel width is largest when geometric expansion is maximal, effectively compensating for particle separation while preserving sharpness at the endpoints.

However, we observe that the adaptive approach fails to work well on outliers: the time-varying kernel width  $\sigma(t)$  becomes too

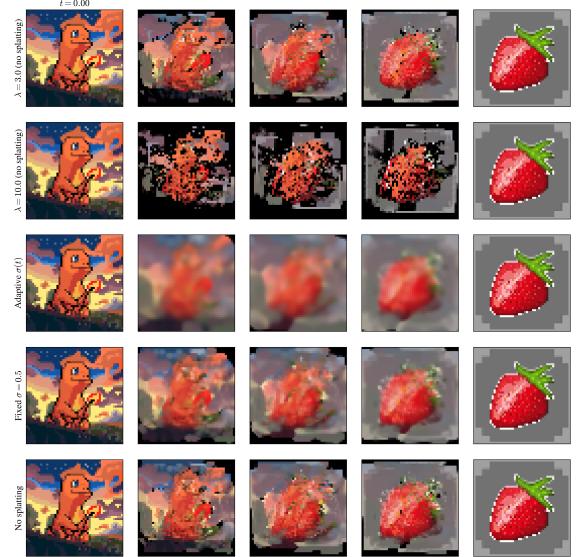
large at intermediate times, causing excessive smoothing that removes outlier pixels that are characteristic of Pixel Art. In contrast, a fixed kernel width provides an optimal compromise. For a given image resolution, we set  $\sigma = 0.5 \times \text{resolution}$  (in pixel units), which is equivalent to  $\sigma = 0.5$  for normalized coordinates in  $[0, 1]$ . This fixed value is sufficiently large to remove tearing artifacts by filling gaps between particles in the 2D image plane, yet small enough to preserve the discrete, quantized nature of Pixel Art, including its characteristic outliers. In general, for any image resolution, we use  $\sigma = 0.5 \times \text{resolution}$  as the standard choice, which provides the perfect balance for removing tearing while keeping outliers, making it the preferred approach for Pixel Art morphing where preserving the discrete color palette and sharp boundaries is essential.

**How Splatting Reduces Tearing.** The tearing artifact arises from the fundamental mismatch between Lagrangian particle evolution and Eulerian grid reconstruction. When discrete particles are transported via the optimal transport map  $\phi_t$ , they move as Lagrangian tracers. In regions of geometric expansion ( $|\det \nabla \phi_t| > 1$ ), these particles spread apart, creating gaps in the fixed Eulerian grid. Without splatting, we simply place each particle at its nearest grid point, leaving empty pixels (holes) where the particle density is insufficient to cover the grid. This violates the Nyquist-Shannon sampling theorem: the effective sampling rate becomes too low relative to the grid resolution.

Gaussian splatting addresses this by performing a **Kernel Density Estimation (KDE)** reconstruction in the 2D image plane. After transporting particles in 5D space, we project them back to their 2D spatial coordinates  $(x_t, y_t)$  and reconstruct the image by convolving each particle with a 2D Gaussian kernel. This smoothing operation fills the gaps between particles, ensuring continuous coverage of the grid. The choice of kernel width  $\sigma$  is guided by the Nyquist-Shannon sampling theorem: to avoid aliasing,  $\sigma$  must be at least half the average inter-particle spacing. For normalized coordinates in  $[0, 1]$ , this yields  $\sigma = 0.5$  (or  $\sigma = 0.5 \times \text{resolution}$  in pixel units), which provides the optimal balance: it is sufficiently large to remove tearing artifacts by filling gaps between particles, yet small enough to preserve the discrete, quantized nature of Pixel Art, including its characteristic outliers. While an adaptive approach (Equation (4)) can theoretically minimize the Sinkhorn divergence by adapting to local sampling density, it suffers from excessive smoothing at intermediate times that removes outlier pixels. The fixed value  $\sigma = 0.5 \times \text{resolution}$  therefore represents the perfect compromise for Pixel Art morphing in general: it satisfies the Nyquist-Shannon condition while preserving the visual characteristics that define this art form. Figure 7 illustrates the reduction of tearing artifacts: without splatting (rows 1-2), severe tearing occurs at  $t = 0.5$ ; adaptive splatting (row 3) reduces tearing but removes outlier pixels; fixed splatting with  $\sigma = 0.5 \times \text{resolution}$  (row 4) provides the optimal balance, effectively removing tearing while preserving discrete characteristics.

## 6 DISCUSSION AND LIMITATIONS

**Scalability vs. Interpretability.** Our explicit reconstruction of  $\pi \in \mathbb{R}^{N \times N}$  provides access to microscopic transport behavior, enabling our density-aware splatting. However, this restricts us to resolutions where  $N^2$  fits in GPU memory. Implicit methods using



**Figure 7: Comparison of adaptive vs fixed vs no splatting to illustrate tearing reduction.**

symbolic tensors (KeOps) allow scaling to megapixels but treat the plan as a "lazy tensor," making local density estimation significantly harder to retrieve.

**Theoretical Convergence.** While we provide strong empirical evidence that our adaptive smoothing minimizes  $S_\epsilon$ , a formal proof is missing. Proving that the sequence of measures generated by our variable-bandwidth KDE converges to the theoretical geodesic as  $N \rightarrow \infty$  would require complex developments in the theory of approximation of measures, likely involving non-asymptotic bounds on the empirical Sinkhorn divergence [8].

**Future Work:** Our 5D lifting relies on Euclidean distance in RGB. However, the RGB cube is not perceptually uniform. A geodesic path between two colors in RGB space does not necessarily correspond to a linear perceptual transition for the human eye. A more rigorous approach would involve lifting the images into a perceptually uniform space like CIELAB or CIELUV. Furthermore, the global hyperparameter  $\lambda$  assumes a constant trade-off between spatial and chromatic costs. A local metric learning approach, using a spatially varying Mahalanobis distance  $C((x, c), (y, d)) = \|x - y\|^2 + (c - d)^T M(c - d)$ , could better respect the local texture manifolds. Even with an optimal transport plan, the interpolated color  $c(t) = (1 - t)c_s + tc_t$  might traverse "muddy" or desaturated regions of the color gamut.

## 7 CONCLUSION

This work addresses fundamental limitations in discrete optimal transport for image morphing, particularly when bridging images with disjoint color distributions and distinct intrinsic resolutions. We identified three critical artifacts that plague standard approaches: *ghosting* from marginal channel processing, *outlier sensitivity* from strict mass conservation, and *tearing* from the Lagrangian-Eulerian discretization gap.

Our contributions provide a unified, mathematically grounded solution. First, we formulated the problem as **5D Joint Unbalanced Optimal Transport**, lifting images into the product space  $\mathbb{R}^2 \times \mathbb{R}^3$  to enforce chromatic coherence while allowing adaptive mass creation/destruction through Csiszár divergence penalties. This approach eliminates ghosting artifacts by transporting mass based on feature similarity (spatial position and color) rather than spatial proximity alone. Second, we introduced **Gaussian Splatting** reconstruction to address the tearing phenomenon. A fundamental aspect of our approach is the clear separation between transport matching and reconstruction: while the optimal transport matching is performed in 5D space to respect the joint space-color geometry, the reconstruction step projects the transported particles back to the 2D image plane and applies a 2D Gaussian kernel for density estimation. This two-phase design ensures that chromatic coherence is maintained during transport while geometric tearing is addressed through 2D smoothing. While we derived an adaptive time-varying kernel bandwidth  $\sigma(t)$  that adapts to intrinsic spectral characteristics and geometric expansion, we found that a fixed kernel width  $\sigma = 0.5 \times \text{resolution}$  (in pixel units) provides the optimal balance in general for any image resolution: it effectively removes tearing artifacts in the 2D image plane while preserving the discrete, quantized characteristics of Pixel Art, including outlier pixels that define this art form.

Third, we established **principled hyperparameter selection guidelines** that enable a priori tuning without expensive grid searches. For the chromatic consistency parameter  $\lambda$ , we empirically determined that  $\lambda \approx 2.0$  provides the optimal trade-off between spatial coherence and chromatic fidelity. For the entropic regularization parameter  $\varepsilon$ , we derived both theoretical bounds (based on Nyquist-Shannon sampling) and empirical heuristics (1% of cost scale), identifying  $\varepsilon = 0.01$  as a practical compromise that balances computational efficiency with visual quality. Most importantly, for the reach parameter  $\rho$  in unbalanced OT, we developed a worst-case uniform distribution heuristic that enables a priori selection: by assuming the balanced transport plan acts uniformly distributed, we compute a conservative bound  $\rho > \frac{1}{2}\mathbb{E}[\|x - y\|^2]$  that ensures geometric transport is favored for typical displacements. This heuristic is validated empirically, with the optimal  $\rho$  value occurring near the theoretical prediction.

Empirically, we demonstrated that fixed splatting with  $\sigma = 0.5 \times \text{resolution}$  provides the optimal compromise for Pixel Art morphing in general: while adaptive splatting can theoretically minimize the Sinkhorn divergence, it suffers from excessive smoothing that removes outlier pixels. The fixed approach effectively removes tearing in the 2D image plane through Gaussian smoothing while preserving the visual characteristics that define Pixel Art. The key insight is that by performing transport matching in 5D space and reconstruction in 2D space, we can simultaneously respect color-space geometry and address geometric tearing. These principled hyperparameter guidelines enable robust transport in outlier-prone environments like Pixel Art, where discrete color palettes and sharp boundaries make images highly sensitive to spurious noise patterns.

Our work bridges the gap between the elegant continuous theory of optimal transport and its practical discrete implementation, demonstrating how a two-phase approach—5D transport matching

followed by 2D Gaussian splatting reconstruction—can effectively address the fundamental challenges of discrete image morphing. This opens new avenues for robust image interpolation in computer vision applications where preserving both geometric and chromatic coherence is essential.

## REFERENCES

- [1] Debarghya Mukherjee, Soham Mukherjee, and Shayan Oveis Gharan. 2021. Outlier-robust optimal transport. In *International Conference on Machine Learning*. PMLR, 7850–7860.
- [2] Kilian Fatras, Thibault Séjourné, Rémi Flamary, and Nicolas Courty. 2021. Unbalanced minibatch optimal transport: applications to domain adaptation. In *International Conference on Machine Learning*. PMLR, 3186–3197.
- [3] Séjourné, T., Feydy, J., Vialard, F. X., Trouvé, A., & Peyré, G. (2019). *Sinkhorn divergences for unbalanced optimal transport*. arXiv preprint arXiv:1910.12958.
- [4] Peyré, G., & Cuturi, M. (2019). *Computational optimal transport: With applications to data science*. Foundations and Trends® in Machine Learning.
- [5] Cuturi, M. (2013). *Sinkhorn distances: Lightspeed computation of optimal transportation distances*. arXiv preprint arXiv:1306.0895.
- [6] Chizat, L., Peyré, G., Schmitzer, B., & Vialard, F. X. (2018). *Scaling algorithms for unbalanced transport problems*. Mathematics of Computation.
- [7] Feydy, J., Séjourné, T., Vialard, F. X., Amari, S. I., Trouvé, A., & Peyré, G. (2019). *Interpolating between optimal transport and MMD using Sinkhorn divergences*. In AISTATS.
- [8] Genevay, A., Chizat, L., Bach, F., Cuturi, M., & Peyré, G. (2019). *Sample complexity of Sinkhorn divergences*. In AISTATS.
- [9] Charlier, B., Feydy, J., Glaunes, J. A., Collin, F. D., & Durif, G. (2021). *Kernel operations on the GPU, with autodiff, without memory overflows*. Journal of Machine Learning Research.