

Color-Space 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 [GitHub repository](#)

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 $\mathcal{X} \times \mathcal{C}$ 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 normalized units ($0.5 \times 1/64$ in our case) 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_ε 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 achieving 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). 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 $\sigma = 0.5 \times$ resolution in normalized units ($0.5 \times 1/64$ in our case) 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 π with marginals exactly α and β , Unbalanced OT relaxes these constraints using Csiszár divergences D_φ . 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, $\rho > 0$ is the marginal penalty parameter that 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 ($D_\varphi = KL$).

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 ε 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:** Under appropriate conditions (compact domain, bounded cost), $S_{\varepsilon, \rho}$ metrizes weak convergence as $\varepsilon \rightarrow 0$. For fixed $\varepsilon > 0$, it metrizes a weaker topology. This implies that a sequence of interpolated measures μ_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: $\alpha = \sum w_i \delta_{(x_i, c_i)}$, where spatial coordinates $x_i \in [0, 1]^2$ are normalized to the unit square and color coordinates $c_i \in [0, 1]^3$ are normalized RGB values. 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 λ acts as a scale multiplier governing the trade-off between spatial displacement and chromatic fidelity. Since both spatial coordinates and color coordinates are normalized to $[0, 1]$, λ explicitly accounts for the scale mismatch between spatial and chromatic distances. We expect λ to be of order 1, which is indeed the case in our experiments ($\lambda \approx 2.0$). A high λ penalizes changing color during transport, favoring "teleportation" (mass destruction/creation via the Unbalanced mechanism) over inconsistent color blending.

The theoretical and practical properties of unbalanced optimal transport discussed above naturally raise the critical question: **How should we choose the key hyperparameters ρ and ε in our 5D color transport applications?**

Several insights from the literature guide our approach. As noted by [8]:

"If $\sqrt{\rho}$ is too large, the model is highly sensitive to segmentation errors in the input data. The theory of unbalanced OT is thus required to make the model robust to outliers."

This emphasizes that ρ needs to be selected to balance robustness against outliers without making the model overly sensitive to small errors or noise in the data. When ρ is too large, the transport effectively becomes balanced and loses this robustness. When ρ is too small, excessive mass creation and destruction may wash out meaningful structures. In our case, since image segmentations and color fields typically contain minor, unpredictable artifacts, ρ must be chosen to reflect the desired tolerance to outlier pixels or mislabeled segments.

Similarly, the entropic regularization parameter ε governs the minimal scale of smoothing. As further remarked in [8]:

"The median error is a function of the entropic blur $\sqrt{\varepsilon}$ that prevents overfitting to the target point cloud."

Thus, ε should be set to balance visual sharpness (favoring a small ε) against the need to avoid overfitting to noise and spurious details (favoring a larger ε). Especially in stylized or quantized imagery, the right choice of ε actively shapes the transition between sharp structure and excessive smoothing.

Research question : In our experiments, we approach hyperparameter selection as an interplay between robustness (controlled by ρ) and sharpness (controlled by ε). For each dataset and morphing scenario, we systematically adjust $\sqrt{\rho}$ and $\sqrt{\varepsilon}$, guided by domain knowledge about typical noise/errors and by qualitative visual analysis of morphs. Open questions remain about optimal

selection strategies, which we address empirically in Section ?? through targeted ablation studies.

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 $K_{ij} = \exp(-C_{ij}/\varepsilon)$. 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 ε 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:

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

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 Geodesic Interpolation in the Joint Space

We interpolate between the source $\alpha = \sum_i w_i \delta_{(x_i, c_i)}$ and target $\beta = \sum_j w_j \delta_{(x_j, c_j)}$ measures in 5D using a simple linear interpolation along the support of the optimal unbalanced transport plan π .

For each pair (i, j) with nonzero mass π_{ij} , we define

$$(x_t^{(ij)}, c_t^{(ij)}) = (1-t)(x_i, c_i) + t(x_j, c_j) \quad (3)$$

and construct

$$\mu_t = \sum_{i,j} \pi_{ij} \delta_{(x_t^{(ij)}, c_t^{(ij)})}. \quad (4)$$

This path preserves the joint geometry of positions and colors, even if the total mass of μ_t varies due to the unbalanced nature of transport. While more rigorous unbalanced geodesics exist [6, 7], this heuristic offers a coherent and computationally efficient interpolation for image morphing.

3.3 The "Debiasing" Trap in GeomLoss

A subtle but critical challenge arises from the regularization conventions used in libraries like GeomLoss [8]. By default, these solvers compute the gradient of the *Sinkhorn Divergence* S_ε , which implies internal "debiasing" steps (subtracting symmetric potentials). However, to reconstruct the specific transport plan π between α and β , we require the dual potentials of the raw entropic cost $OT_\varepsilon(\alpha, \beta)$, 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 Equation (2) to recover the correct mass scaling.

4 HOW TO CHOOSE HYPERPARAMETERS

We evaluate our pipeline on a challenging morphing task: transforming low-resolution pixel-art sprites (Source, 64×64). These images exhibit disjoint supports in the color histogram and topologically distinct features.

4.1 How to Choose λ

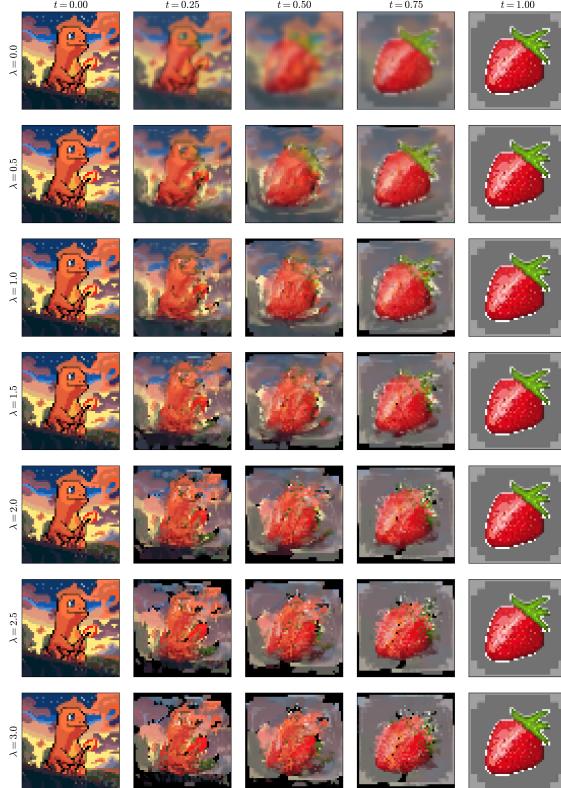


Figure 1: Interpolation timelines for different values of λ .

Figure 2 shows the displacement fields.

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

- **Regime $\lambda \rightarrow 0$ (Pure Spatial Matching):** When $\lambda = 0$, color is ignored and the cost depends only on spatial position: $C = \|x - x'\|^2$. Each pixel is matched solely by its location, regardless of its color, so there is no meaningful color transport. This leads to color values simply blending

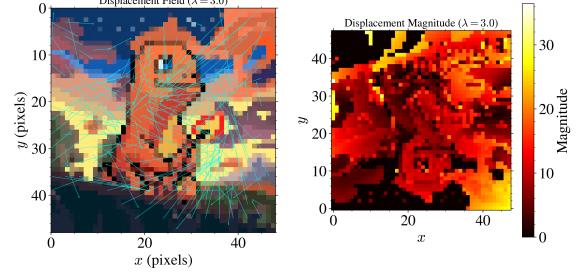


Figure 2: Displacement fields.

or fading into whatever is present at corresponding positions in the target image, yielding smooth but potentially unnatural color transitions rather than real movement of colored features.

- **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 threshold defined by ρ (where transport cost exceeds 2ρ), the Unbalanced mechanism destroys the mass.

Conclusion for λ scale. We expected and confirmed that λ stays of order one due to the normalized 5D transport. 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 λ 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 λ 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 in our specific case a trade-off between spatial coherence and chromatic fidelity.

4.2 How to Choose ϵ a Priori

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

To guide the selection of ϵ , we first determine an appropriate value for ρ 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 ϵ values. This allows us to assess the robustness of the transport quality with respect to the blur parameter when the marginal penalty parameter ρ is set to a value that ensures geometric transport. Figure 3 shows the interpolation timelines for different ϵ 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 ϵ value.

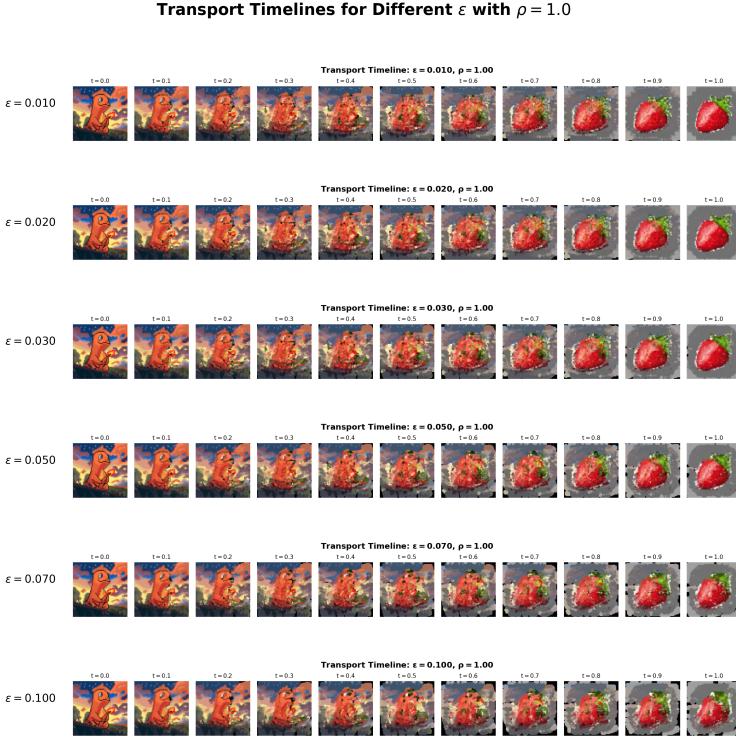


Figure 3: Transport timelines for different values of ε with $\rho = 0.7$ fixed.

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.

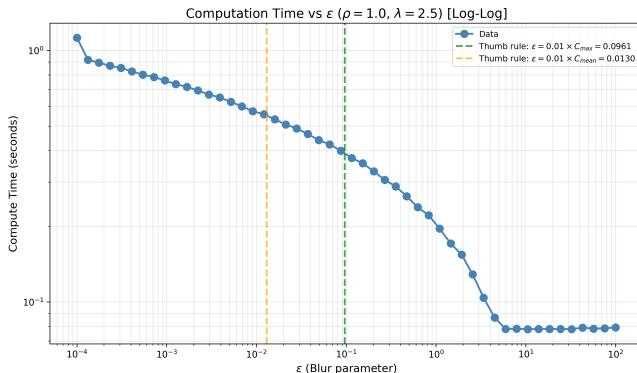


Figure 4: Computation time for transport plan calculation as a function of ε (log-log scale) with $\rho = 0.7$ fixed.

Optimal ε choice To further justify the empirical 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 ε values (from 10^{-4} to 100) with $\rho = 0.7$ fixed. The decreasing variation is expected, as larger ε accelerates Sinkhorn convergence at the cost of increased blur in the transport plan.

Even though we are computing transport in 5d space, the blur radius that leads to visible artifacts comes from the spatial loss. So our analysis focus on the pixel displacement. Theoretically, the minimal ε value should be chosen such that the effective blur scale $\sqrt{\varepsilon}$ matches the typical size of moving pixels [3]. For a 64×64 image, as shown in [?] the pixel typical moving size δ is 10 pixels leading to $\delta = 10/64$ in coordinates units. Setting $\sqrt{\varepsilon} = \delta$ yields $\varepsilon = \delta^2 \approx 0.025$, which represents the lower bound for ε at which blurring begins to appear.

This is a clear accordance with the empirical observation 3 that visual blurring artifacts for the Pixel-art morphed pictures begins to appear at $\varepsilon \approx 0.01$.

Another common heuristic in optimal transport practice suggests choosing ε as a fraction of the cost scale of the balanced case [5, 11]. The thumb rule states that a good initial value for ε is approximately 1% of the maximum or mean cost:

$$\varepsilon \approx 0.01 \times C_{\max} \approx 0.01 \times (D/2)^2 \quad \text{or} \quad \varepsilon \approx 0.01 \times C_{\text{mean}} \quad (5)$$

For the cost scale heuristic, we consider the cost function (Equation (1)), since both spatial and color transport are primary mechanisms for color-shape morphing. While the full cost function is $C((x, c), (x', c')) = \|x - x'\|_2^2 + \lambda^2 \|c - c'\|_2^2$, and we already derived that λ should be of order 1. The derivation accounts for the fact that we operate in a 5D joint space. In the 5D unit hypercube $[0, 1]^2 \times [0, 1]^3$, the maximum distance between two points is $D = \sqrt{5}$ (diagonal of the 5D hypercube). However, we do not expect particles to travel from one corner to the opposite corner; a more realistic maximum displacement is from a corner to the center of the hypercube, which is $D/2 = \sqrt{5}/2 \approx 1.118$. This yields a maximum spatial cost estimate of $C_{\max}^{\text{spatial}} = (D/2)^2 = 5/4 = 1.25$.

Our chosen value $\varepsilon \approx 0.01 \times (D/2)^2$ (with $D = \sqrt{5}$ for the 5D hypercube, this gives $\varepsilon \approx 0.0125$, or $\varepsilon = 0.01$ in absolute terms) falls within this recommended range and satisfies the empirical thumb rule.

conclusion for ε scale. So both approaches lead to the same conclusion: the choice of $\varepsilon \approx 0.01 \times (D/2)^2$ (with $D = \sqrt{5}$, giving $\varepsilon = 0.01$ in absolute terms) represents a well grounded practical compromise.

4.3 How to Choose ρ a Priori for Outliers Environments

Geometric Interpretation of low ρ values behavior. To provide a geometric interpretation of how the mean displacement curve behaves when ρ is small: when ρ is small, we expect a low max displacement. This 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).

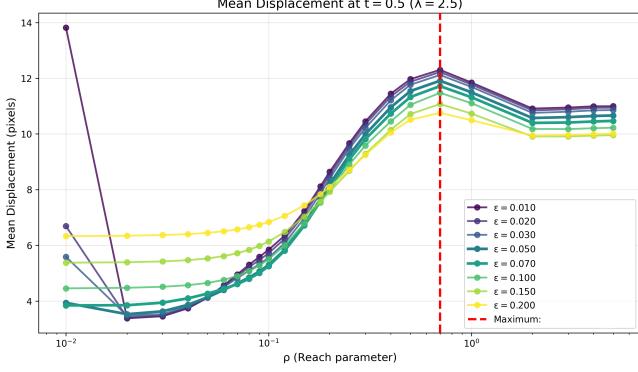


Figure 5: Mean displacement for unbalanced OT reconstructions at $t = 0.5$ as a function of ρ for different values of ϵ ($\lambda = 2.0$).

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 ρ 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 ρ 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 ρ values enforce more geometric transport, resulting in larger but coherent displacements that preserve the shape transformation geometry while maintaining stability.

The goal of this analysis is to enable choosing ρ a priori, without requiring expensive grid searches or trial-and-error.

The marginal penalty parameter ρ defines a characteristic scale for interaction. The threshold 2ρ arises from the optimality conditions of unbalanced OT with KL divergence penalties: when transporting mass from x to y , the solver compares the transport cost $C(x, y)$ with the cost of destroying mass at x and creating it at y , which is 2ρ (one penalty ρ for destruction, one for creation). Transport is favored over creation/destruction only if the transportation cost $\|x - y\|^2 < 2\rho$; otherwise, it is cheaper to locally destroy and create mass. 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 ρ is critical for robust transport.

To ensure that geometric transport is favored for typical displacements, ρ should be chosen such that it exceeds the maximum plausible displacement of the moving pixels [3].

Since we have shown that λ of order one allows have a good color-space balance, we can now derive a simple bound for ρ . Let D denote the diameter of the 5D unit hypercube, so $D = \sqrt{5}$. However,

we do not expect particles to travel the full diameter; a more realistic maximum displacement is from a corner to the center, which is $D/2 = \sqrt{5}/2$. For a displacement of magnitude $D/2$, the transportation cost is $(D/2)^2$. To ensure geometric transport is favored, we require $(D/2)^2 < 2\rho$, yielding:

$$\rho > \frac{(D/2)^2}{2} \quad (6)$$

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

We observe that the maximum of the mean displacement curve occurs near the theoretical prediction given by Equation (6), between 0.5 and 1.0. Figure 5 shows the mean displacement between balanced and unbalanced OT reconstructions at $t = 0.5$ as a function of ρ for different values of ϵ ($\lambda = 2.0$).

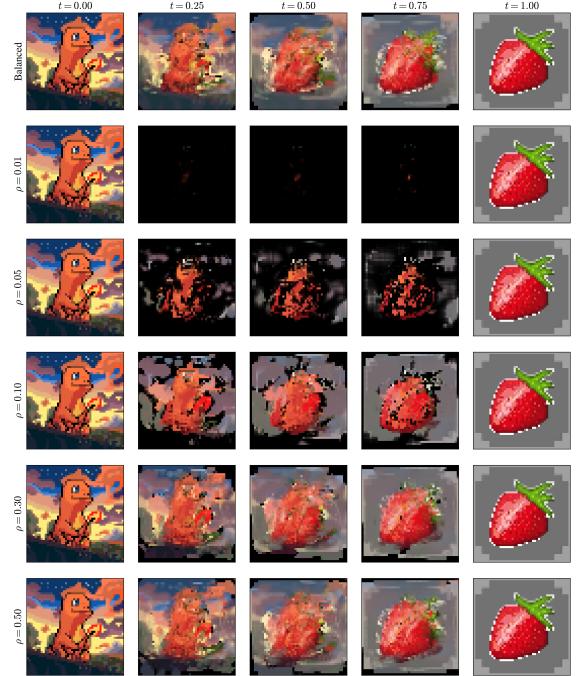


Figure 6: Interpolation timeline for different values of ρ .

From this, we can confirm that unbalanced OT has a desirable influence on the transport map for ρ values satisfying the bound given in Equation (6), which corresponds to values near half the diameter of the 5D hypercube. As a consequence, we can use the heuristic $\rho > \frac{(D/2)^2}{2}$ to select ρ a priori for outlier-prone environments.

5 ADDRESSING THE BLACK-BLOCK TEARING ARTIFACT

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).

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 KDE 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 σ , where σ is the 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. Each transported particle carries its interpolated color $c_t^{(ij)}$ from Equation (3), which is directly assigned to the particle. The Gaussian kernel in the 2D spatial plane then determines how this color is distributed across neighboring pixels, preserving the chromatic coherence established during 5D transport.

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 ϕ_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 filling the gaps between particles, ensuring continuous coverage of the grid. The choice of kernel width σ is guided by the Nyquist-Shannon sampling theorem: to avoid aliasing, σ must be at least half the average inter-particle spacing. Given that spatial coordinates are normalized to $[0, 1]^2$ (as introduced in Section 2), this yields $\sigma = 0.5 \times \text{resolution}$ in normalized units ($0.5 \times 1/64$ in our case). This 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 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 $\sigma = 0.5 \times \text{resolution}$ (in normalized units, $0.5 \times 1/64$ in our case) 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 $\sigma = 0.5 \times \text{resolution}$ in normalized units (row 4) provides the optimal balance, effectively removing tearing while preserving discrete characteristics.

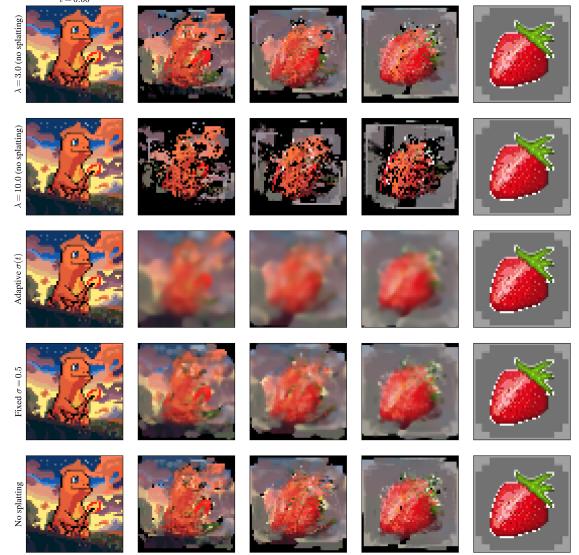


Figure 7: Comparison of adaptive vs fixed vs no splatting to illustrate tearing reduction.

5.1 Hyperparameter Summary

The following summarizes the recommended hyperparameter values, with $D = \sqrt{5}$ for the 5D hypercube:

- $\lambda = O(1)$: Optimal trade-off between spatial coherence and chromatic fidelity. Prevents aberrant hue shifts while allowing slight illumination changes.
- $\epsilon = 0.01 \times (D/2)^2$: Balances computational efficiency with visual quality. Matches theoretical bound $\sqrt{\epsilon} \approx \delta$ and falls within 1% of cost scale heuristic (using corner-to-center distance $D/2$).
- $\rho > \frac{(D/2)^2}{2}$: Ensures geometric transport over mass creation/destruction. For 5D hypercube with diameter $D = \sqrt{5}$, yields $\rho > 0.625$. In practice, $\rho = 0.7$ provides optimal results.
- $\sigma = 0.5 \times \text{resolution}$: Fixed kernel width in normalized units ($0.5 \times 1/64$ in our case) satisfying Nyquist-Shannon condition. Preserves Pixel Art characteristics while reducing tearing.

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 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_ϵ , 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 [9].

Future Work: Our 5D lifting relies on Euclidean distance in RGB. However, the RGB cube is not perceptually uniform. While our geodesic interpolation in the joint space (Equation (3)) correctly follows straight-line paths in RGB space, these paths do not necessarily correspond to perceptually uniform transitions for the human eye. A geodesic between two colors in RGB space may traverse "muddy" or desaturated regions of the color gamut, even though it minimizes the Euclidean distance. A more rigorous approach would involve lifting the images into a perceptually meaningful color space, such as CIE XYZ, which forms the foundation for colorimetry. The CIE XYZ space is designed so that linear combinations of its coordinates correspond to physically realizable colors, and distances in this space provide a more meaningful measure of perceptual difference than RGB. While further transformations to CIELAB or CIELUV yield spaces with near-perceptual uniformity, even lifting to CIE XYZ improves the color interpolation paths and offers better alignment with how humans perceive color transitions. Additionally, our method currently uses a global hyperparameter λ to trade off spatial and chromatic costs uniformly across the image; exploring local adaptation of this trade-off in future work could further improve the perceptual quality of the results.

7 CONCLUSION

We addressed key limitations in discrete optimal transport for image morphing, especially for images with disjoint color distributions and different resolutions. Standard optimal transport methods often suffer from three main artifacts: *ghosting*, *outlier sensitivity*, and *tearing*.

To solve this, we formulated the problem as **5D joint unbalanced optimal transport**, which lifts images to $\mathbb{R}^2 \times \mathbb{R}^3$ and allows adaptive mass change using an Unbalanced Sinkhorn divergence. This approach eliminates ghosting by considering both spatial position and color similarity. To reduce tearing, we introduced a two-phase method: optimal transport is computed in 5D, but reconstruction is done via 2D Gaussian splatting, which smooths geometric artifacts while preserving the quantized look of pixel art. We found that a simple, fixed kernel width $\sigma = 0.5 \times \text{resolution}$ works robustly across images.

We also provided practical hyperparameter rules: $\lambda \approx 2.0$ balances spatial and color fidelity, $\varepsilon \approx 0.01 \times (D/2)^2$ ensures quality and efficiency, and $\rho > (D/2)^2/2$ (with $D = \sqrt{5}$) favors transport over mass creation. Our empirical results confirm that these choices yield robust morphing for pixel art, handling noise and outliers effectively.

In summary, our two-phase 5D/2D approach makes discrete optimal transport practical for challenging image morphing tasks, combining theoretical guarantees with settings that work out-of-the-box.

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] Liero, M., Mielke, A., & Savaré, G. (2018). *Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures*. Inventiones Mathematicae, 211(3), 969–1117.
- [8] 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.
- [9] Genevay, A., Chizat, L., Bach, F., Cuturi, M., & Peyré, G. (2019). *Sample complexity of Sinkhorn divergences*. In AISTATS.
- [10] 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.
- [11] Schmitzer, B. (2019). *Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems*. arXiv preprint arXiv:1610.06519.