G-IR: Geometric Image Representation for Learning

Xin Chen[*1], **Qi Zhao**[*1], **Wei Zeng**[1†], **Zongben Xu**[1]

[1]School of Mathematics and Statistics, Xi'an Jiaotong University, China
cxincx@stu.xjtu.edu.cn, shmilyqi@stu.xjtu.edu.cn, wz@xjtu.edu.cn, zbxu@xjtu.edu.cn

Abstract

Images are generally represented by pixel intensities or color values, which are usually used as direct inputs for learning. This study innovatively proposes a geometric image representation method and refreshes the general learning model (e.g., autoencoder) in the diffeomorphic space. Based on the theory of geometric optimal transport and quasiconformal mapping, we equivalently transform the intensity representation into a shape representation. The image space becomes a diffeomorphic space, where any image can be uniquely represented as a Beltrami coefficient function defined on a uniform grid reference, and vice versa. This innovative geometric image representation (G-IR) captures the fine-grained structure inherent in the entire image, which is different from the traditional feature extraction that focuses on the internal geometric objects of the image (such as boundaries and axes). The diffeomorphic property preserves structure in the generation process, which is very necessary in the field of real physics. It can be assembled into existing pipelines as a plug-in, providing structure-preserving properties for the entire framework. Experiments on image restoration and interpolation validated the high efficiency, efficacy and applicability of the G-IR method, demonstrating its superior performance compared to common pixel-level image appearance representations.

Introduction

As the most widely used data set, images are the first focus of artificial intelligence research and deep learning models. The essential representation of images is the core foundation, which affects the feature extraction of data and the performance of learning models in various tasks.

Traditional representations focus either on local information, such as the widely used pixel intensity (Sisodia and Verma 2011), or on global information, such as the temporal and frequency information of Fourier transform on pixel intensities (Narwaria et al. 2012). Pixel-level independent information cannot capture global structures, so some techniques are designed to make up for this representation deficiency, such as using convolutional kernels or multiple layers to expand the field of view and obtain deeper and broader understanding. At the same time, pixel-level statistics lose fine-grained values and internal structures.

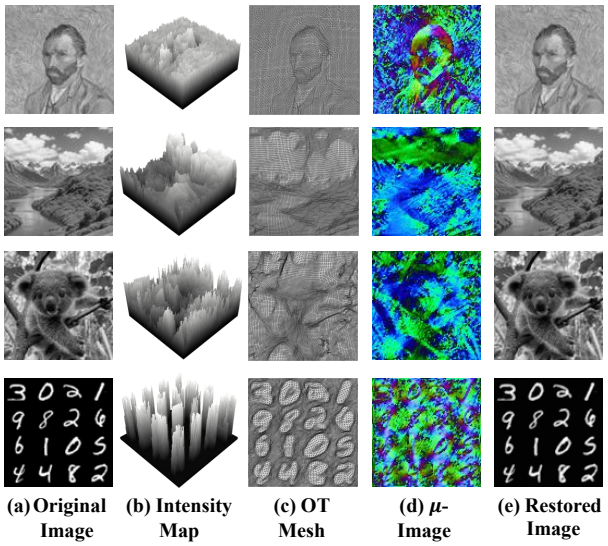
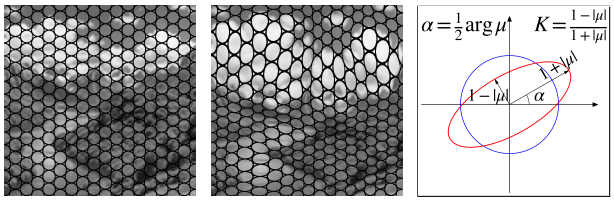


Figure 1: Overview of our G-IR: An input image's intensity (a) is conceptualized as a height field (b), which defines a target measure for optimal transport. This process yields an intensity-aware mesh (c), whose underlying deformation is encoded as a Beltrami coefficient (μ) (d). This G-IR, in turn, enables high-fidelity quasiconformal restoration (e) of (a).

Given the characteristics of existing image representation methods, we began to think about using modern geometry perspectives to explore more intrinsic structures in images, expecting to obtain their geometric structure based on their appearance. We utilize the theories of optimal transport (OT) and quasiconformal (QC) geometry to convert an image to a geometric mapping representation. Figure 1 shows few examples on various gray-scale images to illustrate our intuition. We first take the intensity as the mass of each pixel, imaging it as a pile of sand at each pixel. The entire image now looks like a pile of sand at varying heights within the region. We then use an optimal transport procedure to flatten the piles to the same height. The mass is then converted to be the area of the sand associated with the original pixel. Geometrically, optimal transport provides a mapping

[*]These authors contributed equally.
[†]Corresponding author.

from the uniform pixel grid to an irregular grid (a triangular mesh is required in computation). This mapping is formulated as a quasiconformal mapping and represented as a complex-valued Beltrami coefficient (BC) function. The Beltrami coefficient describes the angle distortion from an infinitesimal circle to an infinitesimal ellipse (see Figure 2), implying eccentricity K and argument α. This transformation is guaranteed to exist, be unique, and be reversible. Furthermore, the entire mapping process is guaranteed to be diffeomorphic, meaning that there are no orientation flips or self-intersections. Therefore, the image is equivalently converted to a geometric representation (with the accompanying total mass) with no information loss, which is theoretically guaranteed and numerically verified. Unlike recent implicit neural representations (INRs) (Dupont et al. 2022; Sitzmann et al. 2020) or 2D Gaussian encodings (Zhang et al. 2024; Zhu et al. 2025), our proposed G-IR deterministically computes a geometrically interpretable Beltrami coefficient, which is firmly grounded in optimal transport and quasiconformal theories, ensuring a precise bidirectional mapping.



(a) Original (b) Deformed (c) BC μ

Figure 2: Illustration of QC map: $\mu = |\mu|e^{i\theta}, \theta = \arg\mu$.

Novelty and Advantages

The proposed geometric image representation (G-IR) constructs new ways to define, compute, and analyze the internal structure of images, mathematically transforming the discrete pixel-wise appearance into a cooperative diffeomorphic geometric structure. To the best of our knowledge, this is the first study on image geometry and the idea of using geometric mapping to characterize images and thus extract the structural geometric information of images.

Operating on this structured representation naturally preserves the structure of the image, so that the smoothness and directionality of the image appearance are not destroyed. This particular property is highly desirable in practice, especially in learning tasks such as image evolution, registration, and generation. In addition, the encoding and decoding transformation modules of G-IR can be used as plug-ins, located after the initial input and before the final output, respectively, and can be used alone or in combination with image pixel appearance for representation learning.

Contributions

Briefly, this work makes the following contributions:

- It innovates image representation in a unique geometric viewpoint, studying the image geometry and converting image appearance to image shape for the first time. This integrates mass uniformization by optimal transport map and shape representation by quasiconformal mapping.
- It creates a method that naturally preserves structure in image deformation-related operations.

The superior performance of G-IR has been validated through the experiments on image restoration and interpolation tasks both qualitatively and quantitatively.

Related Work

Here we review the most related works in terms of both methodology and computation.

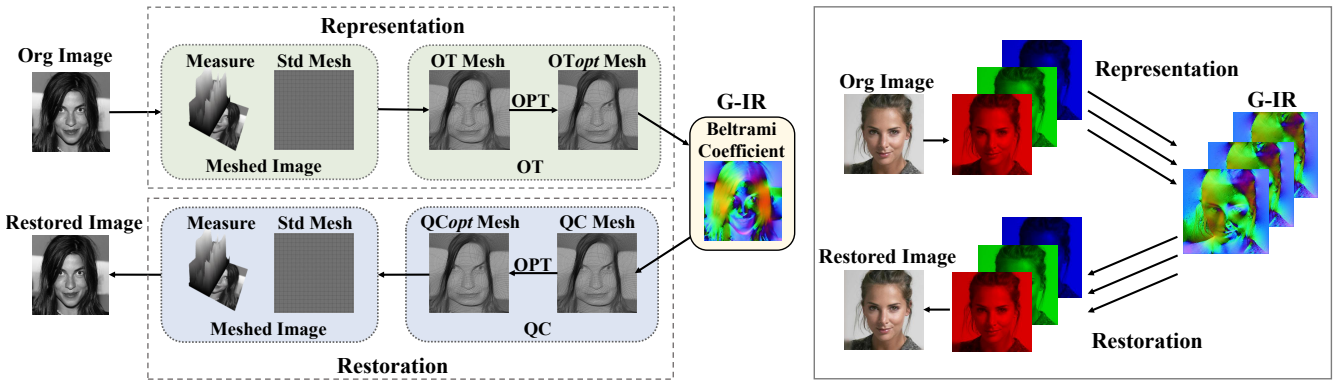
Image Representation

The representation of an image is fundamental to computer vision, with methodologies evolving from hand-crafted features to deep learned representations. Early work focused on extracting explicit yet often sparse geometric primitives like edges and contours (Djilali et al. 2021; Ai et al. 2023) or local texture descriptors such as SIFT and LBP (Ai et al. 2024; Akimoto et al. 2019). Subsequently, transform-domain methods, including Fourier and wavelet analysis (Akimoto, Matsuo, and Aoki 2022; Albanis et al. 2021), alongside sparse representations over learned dictionaries (Apitzsch, Seidel, and Hirtz 2018; Armeni et al. 2017), gained prominence by encoding images through basis coefficients. While powerful, these classic approaches typically struggle to capture dense, holistic image structure in a geometrically explicit manner. In the modern era, deep learning, particularly with CNNs and ViTs (He et al. 2016; Dosovitskiy et al. 2020; Artizzu et al. 2021), has become the dominant paradigm for learning powerful, hierarchical features directly from raw pixels. A fundamental limitation, however, is that these learned features remain deeply entangled with photometric properties (Bai et al. 2024; Ban et al. 2020), treating intrinsic structure as an implicit byproduct rather than a primary, controllable entity.

Our work addresses this gap by proposing a framework that reformulates the entire image as a diffeomorphic deformation field, where the intrinsic geometry itself serves as a complete and robust representation of the image structure.

Geometric Mapping

In recent decades, two powerful theoretical tools, computational conformal geometry (Gu, Luo, and Yau 2010) and computational optimal transport (Peyré, Cuturi et al. 2019), have emerged and achieved wide successes in practice. They provide essential tools to compute geometric mappings, obtaining conformal mapping (angle-preserving) and optimal transport map (measure-preserving), respectively, both of which can be subsumed by quasiconformal mapping. Here, geometric mapping denotes transforming a geometric shape to another, preferably a diffeomorphism (one-to-one, onto). QC mapping, with roots in the classical work of Ahlfors and Bers (Ahlfors 2006), uses the Beltrami coefficient to control or describe local angle distortion of a diffeomorphism (Zeng et al. 2009). Conformal mapping (Gu and Yau 2003) is a special case of QC mapping, where angle distortions are zero.



(a) Pipeline (Single-Channel) **(b) Multi-Channel Case**

Figure 3: The overall pipeline for our proposed G-IR. Illustration (a) details the G-IR generation and restoration for a single-channel image. The target measure from pixel intensities guides an Optimal Transport (OT) map from a standard uniform mesh into a non-uniform OT mesh. The local deformation yields the Beltrami coefficient (BC) field, our G-IR. For restoration, the BC determines a unique quasiconformal (QC) map of a standard mesh and obtains the QC mesh for pixel intensity recovery. For a multi-channel image (b), each channel is operated independently and finally combined to restore the image.

One efficient method is to feed the auxiliary metric associated with Beltrami coefficients (Zeng et al. 2012) to conformal mapping algorithms, such as holomorphic 1-forms (Zeng, Lui, and Gu 2014) and curvature flows (Zeng and Gu 2013), to achieve QC mapping. In parallel, originating from Monge's problem (Monge 1781). OT seeks a cost-minimal transport map from source distribution to target distribution (Kusner et al. 2015; Huang et al. 2016; Bonneel et al. 2011; Solomon et al. 2015). Geometric semi-discrete OT map (An et al. 2021) was proposed by computing gradient of Brenier potential, along with theoretical guarantee of uniqueness and existence (Gu et al. 2013), which works for discrete points from a specific density distribution to a uniform one.

Our work extends the semi-discrete OT map to triangular meshes via minimizing an energy functional, thereby optimizing the OT map while preserving the mesh topology (i.e., triangle connectivity). The optimized OT map is then formulated as a quasiconformal mapping.

Method

The overall pipeline of our proposed framework is designed to be an end-to-end differentiable process, which consists of two opposing stages: image representation and image restoration, as shown in Figure 3. It transforms the image's intensity distribution to a geometric distortion representation on a background standard mesh (see Algorithm 1), then uses this distortion to restore the input image (see Algorithm 2). This transformation is achieved by synergizing optimal transport map and quasiconformal map. G-IR is discussed for grayscale images here; by independently processing each RGB channel, it can be easily extended to color images.

Image to Measured Mesh

Given an image, we need to convert it into a triangular mesh, associated with the image intensity function as the target density function on it. The pixel intensity is considered as a mass. All pixel masses are to be smoothened collaboratively within the mesh domain by optimal transport.

We first initialize a **Standard Mesh**, denoted as $\mathcal{M}_{std}$, on the image domain $\Omega = [0, W] \times [0, H]$. The vertices V_{std} of the mesh are located at the centers of the image pixels, and its topological structure is defined by triangulating the square grid constructed on these vertices, specifically by adding diagonal edges in the same direction. This regular mesh $\mathcal{M}_{std}$ serves as the canonical reference domain for our subsequent geometric transformations.

Then, we interpret the normalized grayscale image as a

Algorithm 1: Image to G-IR

Input: Image I with grayscale intensities g
Parameter: $\epsilon, \lambda_u, \lambda_r$
Output: G-IR μ and $\mathcal{G}$

Initialize standard mesh $\mathcal{M}_{\text{std}}$ with vertices V_{std}
$\mathcal{G} \leftarrow \sum_j (g_j + \epsilon)$
$s_i^t \leftarrow (g_i + \epsilon)/\mathcal{G}$
$V_{\text{OT}} \leftarrow \text{OT\_Solve}(\mathcal{M}_{\text{std}}, \{s_i^t\})$
$V_{\text{OT}}^* \leftarrow \arg\min_V E_g(V)$ from Equation (3)
$\mu \leftarrow \mu(V_{\text{std}}, V_{\text{OT}}^*)$
return $\mu, \mathcal{G}$

Algorithm 2: Image restoration from G-IR

Input: G-IR μ, $\mathcal{G}$, ϵ
Parameter: γ_1, γ_2
Output: Restored image I

$V_{\text{QC}} \leftarrow \text{QC\_Solve}(\mu)$
$V_{\text{QC}}^* \leftarrow \arg\min_v E_\mu(v)$ from Equation (5)
$\alpha_i \leftarrow s_i(V_{\text{QC}}^*)/\sum_j s_j(V_{\text{QC}}^*)$
$g_i^t \leftarrow \mathcal{G}\alpha_i - \epsilon$.
return $I(g^t)$

scalar intensity function $g : \Omega \to [0, 1]$, where Ω denotes the background continuous domain. The standard mesh $\mathcal{M}_{\text{std}}$ is naturally endowed with a uniform source measure, which we define as the normalized Lebesgue measure on Ω, denoted by m_S. This represents the initial uniform density of the geometric space. At the same time, we define the target measure m_T by defining a target density function $\rho_T : \Omega \to \mathbb{R}^+$ derived directly from the intensity function g.

Specifically, we define the target density ρ_T to be proportional to the image intensity. To handle regions of zero intensity, a small positive hyperparameter ϵ is introduced. The continuous target density is then formulated as the normalized, perturbed intensity:

$$\rho_T(x) = \frac{g(x) + \epsilon}{\int_\Omega (g(y) + \epsilon) dy} \tag{1}$$

The denominator serves as a normalization factor, ensuring that the resulting target measure m_T is a probability measure (i.e., $\int_\Omega \rho_T(x)dx = 1$). The value of ϵ can be tuned according to the dataset's characteristics.

The target measure is thus given by $dm_T(x) = \rho_T(x)dm_S(x)$. In the discrete setting of our mesh, this corresponds to the per-vertex target area s_i^t for vertex v_i. Since the source measure is uniform, we can set the target measure for each vertex to be its normalized intensity value:

$$s_i^t = \hat{g}_i \quad \text{where} \quad \hat{g}_i = \frac{g_i + \epsilon}{\sum_j (g_j + \epsilon)} \tag{2}$$

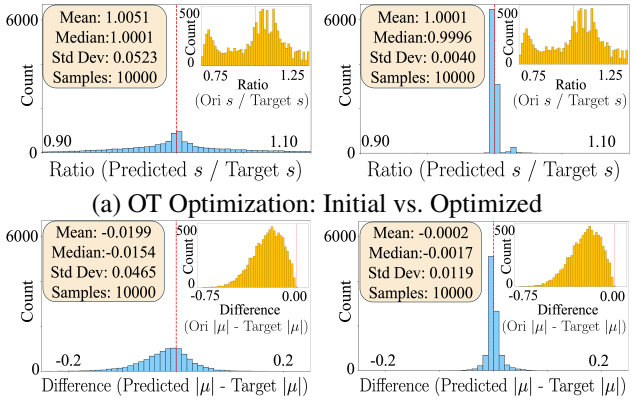
We record the sum of all perturbed intensity values as the total intensity $\mathcal{G} = \sum_j (g_j + \epsilon)$, which will be utilized in the subsequent image restoration phase.

Optimal Transport Map of Image

With the standard mesh $\mathcal{M}_{\text{std}}$ established along with both the source measure m_S and target measure m_T, our objective is to warp it from the uniform source measure m_S into a configuration with the intensity-aware target measure m_T. This is achieved by finding an optimal transport map $T : \Omega \to \Omega$. This map satisfies the measure preservation constraint.

We apply the geometric semi-discrete optimal transport solver (An et al. 2021) as an integrated component of our framework to achieve the initial OT map. It solves the Monge problem to yield the new vertex positions $V_{\text{OT}} = T(V_{\text{std}})$, which define our initial intensity-aware **OT Mesh** along with the original mesh topology. Because this method inherently operates on discrete points and also has numerical errors due to discrete computation, we subsequently refine the map via minimizing an energy functional to improve the approximation accuracy while preserving the triangular mesh topology.

OT Optimization. For general applications the initial OT map might suffice, but in the context of image restoration, even minor geometric deviations in V_{OT} can propagate through our restoration function $g(\cdot)$ and manifest as obvious visual artifacts. This stage fine-tunes the vertex positions $V = \{v_i\}$ by minimizing an energy functional $E_g(V)$ that is designed to balance direct image-space fidelity with the regularity of the underlying deformation field.



(a) OT Optimization: Initial vs. Optimized

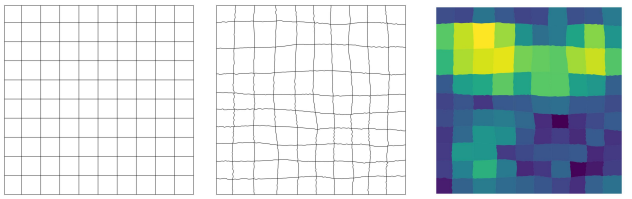
(b) QC Optimization: Initial vs. Optimized

Figure 4: Optimization of maps for the image in Figure 3a.

The energy functional $E_g(V)$ is constructed around our differentiable restoration function $g(V)$. Conceptually, the restoration acts as the inverse of our initial intensity-to-measure transformation. It is a composite function $g = \mathcal{R} \circ \mathcal{A}$, where $\mathcal{A}$ maps vertex positions V to a set of local vertex areas $\{s_i(V)\}$, and $\mathcal{R}$ is the restoration operator that applies the inverse of Equation (2) to these areas to yield pixel intensities. The total energy $E_g(V)$ seamlessly integrates two primary objectives, formulated to capture both global and local error characteristics:

$$\begin{aligned} \arg\min_V E_g(V) = & \underbrace{\frac{1}{|\Omega|}\int_\Omega |g(V)(x) - g_{\text{truth}}(x)|^2 dx}_{\mathcal{L}_{g\text{-fidelity}}} \\ & + \lambda_u \underbrace{\sup_{x\in\Omega} |g(V)(x) - g_{\text{truth}}(x)|}_{\mathcal{L}_{g\text{-uniform}}} \\ & + \lambda_r \underbrace{\sup_i \|v_i - (V_{\text{OT}})_i\|_2}_{\mathcal{L}_{g\text{-reg}}} \\ s.t. \quad & \|\mu(V_{\text{std}}, V)\|_\infty < 1 \end{aligned} \tag{3}$$

The primary component of this functional is the image fidelity objective, which aims to make the restored image $g(V)$ is faithful to the ground-truth g_{truth} . Its integral term $\mathcal{L}_{g\text{-fidelity}}$ minimizes the average restoration error across the domain, while the supremum term $\mathcal{L}_{g\text{-uniform}}$, weighted by λ_u, penalizes the worst-case pointwise deviation, thereby suppressing local artifacts and enforcing uniform quality. This is complemented by the deformation regularizer $\mathcal{L}_{g\text{-reg}}$, weighted by λ_r, which constrains each optimized vertex v_i to a trust region around its initial position $(V_{\text{OT}})_i$. The constraints given by the Beltrami coefficients μ of the map preserve the geometric structure and prevents mesh self-intersections, thereby ensuring the stability and geometric soundness of the optimization. Figure 4a shows the optimization effect for the example image.



(a) Uniform Grid (b) Deformed Grid (c) Restored Image

Figure 5: Illustration of image restoration from a map.

Quasiconformal Map to Represent OT Map

The deformation field from the standard mesh to the optimized OT mesh can be characterized by a quasiconformal mapping such that its local geometric distortion is represented by a unique Beltrami coefficient function. We therefore take this Beltrami coefficient μ as the intrinsic **geometric representation** (signature) of the image, along with the total intensity $\mathcal{G}$ recorded in the initialization phase, i.e.,

$$\text{G-IR} = (\mu, \mathcal{G}). \tag{4}$$

μ-Image Generation. The map is given by $f : \mathcal{M}_{\text{std}} \to \mathcal{M}_{\text{OT}}$, where two meshes have the same topology. The Beltrami coefficient of the map is denoted by $\mu(V_{std}, V_{OT})$. Mathematically, $\mu(z) = f_z/f_{\bar{z}}$, where the condition $|\mu(z)| < 1$ ensures the map is a local, orientation-preserving diffeomorphism. In our discrete framework, we compute a piecewise constant Beltrami coefficient for each triangular face from the corresponding triangles in the standard and OT meshes. The BC for each vertex is computed by averaging the BCs of adjacent triangles. We can represent the Beltrami coefficient field over the domain in an image form, in which the RGB value is set as $(|\mu(z)|, \frac{\sin\theta(z)+1}{2}, \frac{\cos\theta(z)+1}{2})$, where $\mu(z) = |\mu(z)|e^{i\theta(z)}$. Figure 1d shows the examples, which implies the intensity-driven distortions.

We denote this discrete Beltrami coefficient as μ_{OT}, now serving as our geometric signature, the problem is inverted: we seek to restore vertex positions V_{QC} from μ_{OT} on the standard mesh. The theoretical foundation for this inversion is provided by the *Measurable Riemann Mapping Theorem* (Ahlfors 2006), which guarantees the existence and uniqueness of a quasiconformal map for any given Beltrami coefficient field satisfying $\|\mu\|_\infty < 1$. The QC mapping can be achieved by a linear Beltrami solver using the auxiliary metric method (Zeng et al. 2009) for the quadrilateral case.

QC Optimization. The restored V_{QC} from the Beltrami coefficient μ_{OT} provides a geometrically consistent approximation of the deformation. However, this process accumulates errors from two primary sources: the initial discretization in computing μ_{OT} and numerical inaccuracies inherent in the inverse solver. To correct these compounded errors and produce a final mesh with maximum fidelity to the target geometric signature, we introduce a QC refinement process through minimizing an energy functional.

This stage optimizes the vertex positions V by minimizing a composite energy function, $E_\mu(V)$. As this objective is fundamentally about aligning discrete computed quantities, we formulate it directly in terms of vector norms. The structure of this energy is directly analogous to our previous image-space optimization:

$$\begin{aligned}\arg\min_V E_\mu(V) = &\underbrace{\|\mu(V_{\text{std}}, V) - \mu_{\text{OT}}\|_2^2}_{\mathcal{L}_{\mu\text{-fidelity}}} \\ &+ \gamma_1 \underbrace{\|\mu(V_{\text{std}}, V) - \mu_{\text{OT}}\|_\infty}_{\mathcal{L}_{\mu\text{-uniform}}} \\ &+ \gamma_2 \underbrace{\|V - V_{\text{QC}}\|_\infty}_{\mathcal{L}_{\text{spatial-reg}}} \\ s.t. \quad &\|\mu(V_{\text{std}}, V)\|_\infty < 1 \end{aligned} \tag{5}$$

Here, the function $\mu(V\text{std}, v)$ denotes the operation of computing the Beltrami coefficient field from a given set of vertex positions V. Drawing a direct parallel to the energy in Equation (3), the $\mathcal{L}_{\mu\text{-fidelity}}$ and $\mathcal{L}_{\mu\text{-uniform}}$ terms work in tandem to drive the mesh's Beltrami coefficient field to match the target signature μ_{OT} in both an average and a worst-case sense. Similarly, the $\mathcal{L}_{\text{spatial-reg}}$ term constrains the optimization to a stable trust region around the initial solution V_{QC}. The mesh topology constraint is also applied. Figure 4b shows the QC optimization effect for the example image.

Image Restoration from QC Map

Given the geometric image representation μ, we recover the OT mesh by the QC mapping with the QC optimization to enhance the precision and quality. Then, we restore the image from the mesh structure with the obtained final vertex positions V_{fin} to carry out mesh-to-image conversion.

In detail, we calculate the area measure s_i^{fin} for each vertex $v_i \in V_{\text{fin}}$ by performing a weighted summation of the areas of its adjacent triangular faces. Let the vector of these vertex area measures be $\mathbf{s}^{\text{fin}} = (s_1^{\text{fin}}, \ldots, s_n^{\text{fin}})$. The final area proportion vector $\boldsymbol{\alpha}$ is obtained by normalization:

$$\boldsymbol{\alpha} = \frac{\mathbf{s}^{\text{fin}}}{\|\mathbf{s}^{\text{fin}}\|_1}. \tag{6}$$

The restoration process, denoted as $g(V)$, recovers the grayscale image by inverting the forward measure-to-intensity mapping previously defined in Equation (2). The optimal transport approximately preserves the measure distribution. Consequently, the recovered target measure $\hat{s}_i^t$ for a given vertex is determined by its final area proportion α_i, $\hat{s}_i^t = \alpha_i$. By algebraically inverting the forward relationship from Equation (2), with the recorded the total intensity $\mathcal{G}$, we directly solve for the restored grayscale value g_i^t:

$$g_i^t = \mathcal{G}\alpha_i - \epsilon \tag{7}$$

Figure 5 shows the image restoration from a map of a uniform grid. Crucially, every operation in the restoration is differentiable. This ensures that the gradients from a pixel-level restoration loss can flow back to the mesh vertex positions V_{fin}, enabling a robust end-to-end optimization.

Experiments

We conduct the following experiments to evaluate the proposed geometric image representation: (1) compute the accuracy of information preservation to validate the fidelity of

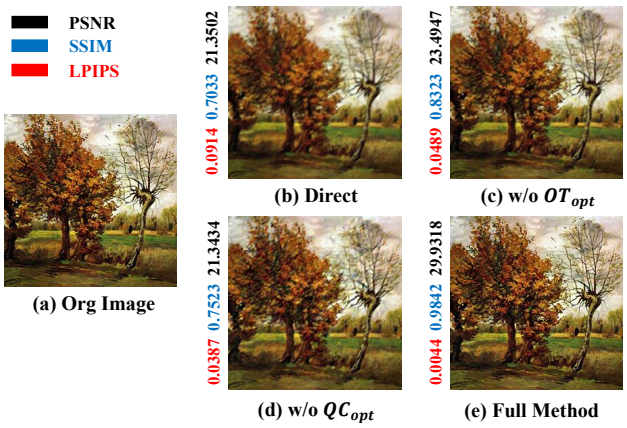


Figure 6: Visualization of ablation study. Our full method (e) most effectively eliminates artifacts and faithfully restores the original image (a), compared to our baseline (b) and variants with missing components (c, d).

Method Variant	PSNR ↑	SSIM ↑	LPIPS ↓
MNIST			
Ours (Full Method)	**33.279**	**0.986**	**0.005**
Ours (Direct)	22.020	0.687	0.298
Ours (w/o OT_{opt})	24.755	0.851	0.201
Ours (w/o QC_{opt})	25.01	0.853	0.195
CIFAR-10			
Ours (Full Method)	**44.451**	**0.9983**	**0.0019**
Ours (Direct)	24.644	0.769	0.284
Ours (w/o OT_{opt})	25.498	0.801	0.227
Ours (w/o QC_{opt})	26.954	0.833	0.161
CelebA-HQ			
Ours (Full Method)	**32.173**	**0.935**	**0.003**
Ours (Direct)	19.871	0.732	0.354
Ours (w/o OT_{opt})	23.605	0.759	0.205
Ours (w/o QC_{opt})	25.136	0.774	0.170

Table 1: Ablation study results on restoration quality. We compare our full method against variants with disabled optimization components. For PSNR and SSIM, higher is better (↑); for LPIPS, lower is better (↓). The best results are highlighted in **bold**, demonstrating the cumulative benefit of our optimization modules.

the image representation pipeline, along with ablation studies on the necessity of the OT and QC optimization modules; and (2) integrate the geometric representation, μ-image, into an autoencoder (AE) architecture, to explore its viability as an effective latent space for generative tasks.

Experimental Setup

Datasets. Our experiments are conducted on three standard datasets: MNIST (LeCun et al. 2002), CIFAR-10 (Krizhevsky and Hinton 2009), and CelebA-HQ (Karras et al. 2017). They cover a wide range of image types, from simple digits and small objects to complex facial structures, allowing for a comprehensive evaluation.

Evaluation Metrics. We assess restoration quality via Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) (Wang et al. 2004), and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018). These metrics evaluate pixel-level, structural, and perceptual fidelity, respectively.

Implementation Settings We implement our framework in PyTorch. The optimization for the stage (Algorithm 1) is performed using the Adam optimizer with a learning rate of 1×10^{-5} over 3000 iterations. The corresponding weights for the energy terms in Equation (3) are set to $\lambda_u = 10.0$ and $\lambda_r = 1.0$. For the stage (see Algorithm 2), we again use Adam but with a learning rate of 1×10^{-4} for 2000 iterations. The weights in the corresponding energy function, Equation (5), are set to $\gamma_1 = 10.0$ and $\gamma_2 = 1.0$. These values provided good results for most cases, but they can be adjusted for specific datasets to achieve optimal performance.All experiments are conducted on a single NVIDIA RTX 4090 GPU.

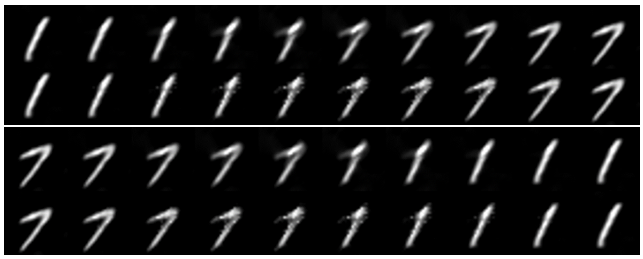
Image Restoration

We perform experiments on the three datasets, each choosing 10 random subsets of 500 test images. Theoretically, the image can be perfectly restored, guaranteed by the Measurable Riemann Mapping Theorem (Ahlfors and Bers 1960; Ahlfors 2006). Despite minor errors from discrete computation, our results demonstrate high fidelity, with SSIM approaching 1.0, LPIPS near 0, and PSNR exceeding 30dB (see Table 1).

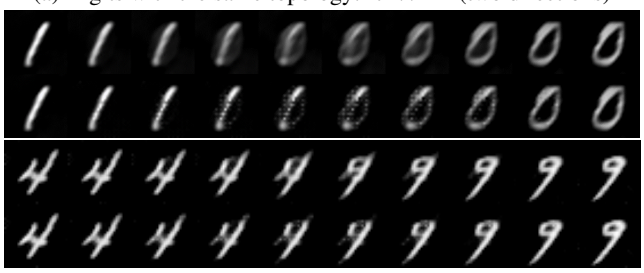
Ablation Study. This study compares our complete framework, **Ours (Full Method)**, against three progressive variants: (1) **Ours (Direct)**, serving as a baseline by disabling both OT and QC optimizations, (2) **Ours (w/o OT_{opt})**, omitting the intensity-driven OT refinement, and (3) **Ours (w/o QC_{opt})**, disabling the QC correction.

Quantitatively, across all datasets, we observe a consistent result that Ours (Direct) version has established a solid baseline and as each optimization module is enabled, the performance metrics improve step by step. This strongly demonstrates that all optimization modules are indispensable to achieving high-fidelity restoration. Of particular note is the significant improvement in the LPIPS metric, especially on the complex CelebA-HQ dataset. This highlights that our method not only excels in improving the original pixel-level accuracy (PSNR), but more importantly, it also excels in preserving the fine structural details necessary for high perceptual quality.

Qualitatively, Figure 6 demonstrates a comprehensive visual comparison. The original image (a) serves as the ground truth. The restoration from the baseline Ours (Direct) (b) captures the coarse structure of the face but suffers from significant geometric distortions and a loss of fine detail. When we disable only the OT optimization, Ours (w/o OT_{opt}) (c)



(a) Digits with the same topology: ‘7’↔ ‘1’ (two directions)



(b) Digits with different topology: ‘1’ ↔ ‘0’, ‘4’ ↔ ‘9’

Figure 7: Image interpolation results by an autoencoder. Top row of each pair: geometric image representation (G-IR); Bottom row: pixel-level image representation (Pixel-IR).

improves the result but still fails to render complex textures, indicating that the OT optimization is crucial. Similarly, disabling only the QC optimization, Ours (w/o QC_{opt}) (d) produces a much clearer image than the baseline, yet subtle artifacts and unnatural smoothness remain in high-frequency areas like hair strands. In stark contrast, **Ours (Full Method)** (e) yields a restoration that is perceptually almost indistinguishable from the original. It successfully corrects the artifacts present in all other variants and faithfully reproduces intricate details and textures. This visual evidence strongly corroborates our quantitative findings and demonstrates that each component in our proposed framework plays an integral role in achieving accurate restoration quality.

Image Interpolation

Having established our representation’s restoration fidelity, we now evaluate its core intrinsic property: the structure preservation and continuity of its induced latent space. To do so, we conduct a latent space interpolation experiment (Kingma and Welling 2013). We compare two models built on an identical U-Net autoencoder architecture (Ronneberger, Fischer, and Brox 2015) : a **Pixel-IR AE**, trained directly on raw image pixels, and our proposed **G-IR AE**, trained on our Geometric Image Representation. For various pairs of digits, we linearly interpolate their corresponding latent vectors and assess the quality of the generated transitions.

As shown in Figure 7, our G-IR AE generates smooth, continuous, and plausible transitions. In contrast, the Pixel-IR AE struggles, showing artifacts like ghosting and noise. Additionally, transitions between intermediate steps may appear abrupt and uneven, with this issue becoming more pronounced when interpolating between digits with different

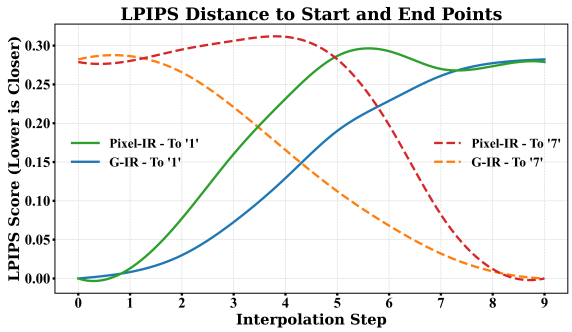


Figure 8: LPIPS distance curves for linear interpolation in the latent space of ‘1’ to ‘7’. The distance-to-endpoints for our method (G-IR) is monotonic, demonstrating a smooth and linear path. The curves for the method (Pixel-IR) fluctuate, indicating a less direct path with inferior continuity.

topological structures (e.g., ‘4‘ to ‘9‘). Figure 8 further confirms this by showing a more direct interpolation path. Quantitatively, our G-IR achieves better adjacent LPIPS scores than Pixel-IR (see Table 2). This verifies that our geometric representation guides the model to learn a latent space that better reflects the intrinsic structure of the data.

Digit Pair	Pixel-IR LPIPS ↓	Ours LPIPS ↓
‘4’ ↔ ‘9’	0.0130	**0.0109**
‘1’ ↔ ‘0’	0.0382	**0.0262**
‘1’ ↔ ‘7’	0.0280	**0.0156**

Table 2: Average adjacent LPIPS scores for latent space interpolation across various digit pairs. Our method consistently achieves lower scores, indicating smoother and more perceptually uniform transitions. The down arrow (↓) indicates that lower is better.

Conclusion

This work establishes a foundation for image representation from a geometric perspective, providing a new approach for intelligent image processing and analysis. It innovatively uses an optimal transport map to homogenize image intensity mass to discover image geometric structure, further representing it as a quasiconformal mapping associated with the Beltrami coefficient function μ. This representation theoretically guarantees the existence, uniqueness, and reversibility. The entire image is then transformed into its intrinsic shape representation, encoded as a μ-image, which can be directly integrated into general models for learning tasks (e.g., autoencoders). The underlying diffeomorphic nature of the transformation means that the operation is performed in a diffeomorphic space, thus preserving the image structure and outperforming existing image representations.

In future work, we will continue to investigate the learning framework in diffeomorphism spaces via G-IR and extend it to a wider range of applications.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China (2021YFA1003002) and the National Natural Science Foundation of China (12090021 and 12090020).

References

Ahlfors, L.; and Bers, L. 1960. Riemann's mapping theorem for variable metrics. *Annals of Mathematics*, 72(2): 385–404.

Ahlfors, L. V. 2006. *Lectures on quasiconformal mappings*, volume 38. American Mathematical Soc.

Ai, H.; Cao, Z.; Cao, Y.-P.; Shan, Y.; and Wang, L. 2023. Hrdfuse: Monocular 360deg depth estimation by collaboratively learning holistic-with-regional depth distributions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13273–13282.

Ai, H.; Cao, Z.; Lu, H.; Chen, C.; Ma, J.; Zhou, P.; Kim, T.-K.; Hui, P.; and Wang, L. 2024. Dream360: Diverse and immersive outdoor virtual scene creation via transformer-based 360 image outpainting. *IEEE transactions on visualization and computer graphics*, 30(5): 2734–2744.

Akimoto, N.; Kasai, S.; Hayashi, M.; and Aoki, Y. 2019. 360-degree image completion by two-stage conditional gans. In *2019 IEEE international conference on image processing (ICIP)*, 4704–4708. IEEE.

Akimoto, N.; Matsuo, Y.; and Aoki, Y. 2022. Diverse plausible 360-degree image outpainting for efficient 3dcg background creation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11441–11450.

Albanis, G.; Zioulis, N.; Drakoulis, P.; Gkitsas, V.; Sterzentsenko, V.; Alvarez, F.; Zarpalas, D.; and Daras, P. 2021. Pano3d: A holistic benchmark and a solid baseline for 360deg depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3727–3737.

An, D.; Lei, N.; Zhao, T.; Si, H.; and Gu, X. 2021. A Moving Mesh Adaption Method By Optimal Transport. In *Proceedings of International Meshing Roundtable*.

Apitzsch, A.; Seidel, R.; and Hirtz, G. 2018. Cubes3d: Neural network based optical flow in omnidirectional image scenes. *arXiv preprint arXiv:1804.09004*.

Armeni, I.; Sax, S.; Zamir, A. R.; and Savarese, S. 2017. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*.

Artizzu, C.-O.; Zhang, H.; Allibert, G.; and Demonceaux, C. 2021. Omniflownet: a perspective neural network adaptation for optical flow estimation in omnidirectional images. In *2020 25th International Conference on Pattern Recognition (ICPR)*, 2657–2662. IEEE.

Bai, J.; Qin, H.; Lai, S.; Guo, J.; and Guo, Y. 2024. GLPanoDepth: Global-to-local panoramic depth estimation. *IEEE Transactions on Image Processing*, 33: 2936–2949.

Ban, Y.; Zhang, Y.; Zhang, H.; Zhang, X.; and Guo, Z. 2020. MA360: Multi-agent deep reinforcement learning based live 360-degree video streaming on edge. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. IEEE Computer Society.

Bonneel, N.; Van De Panne, M.; Paris, S.; and Heidrich, W. 2011. Displacement interpolation using Lagrangian mass transport. In *Proceedings of the 2011 SIGGRAPH Asia conference*, 1–12.

Djilali, Y. A. D.; Krishna, T.; McGuinness, K.; and O'Connor, N. E. 2021. Rethinking 360deg Image Visual Attention Modelling With Unsupervised Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 15414–15424.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Dupont, E.; Kim, H.; Eslami, S.; Rezende, D.; and Rosenbaum, D. 2022. From data to functa: Your data point is a function and you can treat it like one. *arXiv preprint arXiv:2201.12204*.

Gu, D. X.; Luo, F.; and Yau, S.-T. 2010. Fundamentals of computational conformal geometry. *Mathematics in Computer Science*, 4(4): 389.

Gu, X.; Luo, F.; Sun, J.; and Yau, S.-T. 2013. Variational principles for Minkowski type problems, discrete optimal transport, and discrete Monge-Ampere equations. *arXiv preprint arXiv:1302.5472*.

Gu, X.; and Yau, S.-T. 2003. Global conformal surface parameterization. In *Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, SGP '03, 127–137. Goslar, DEU: Eurographics Association. ISBN 1581136870.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Huang, G.; Guo, C.; Kusner, M. J.; Sun, Y.; Sha, F.; and Weinberger, K. Q. 2016. Supervised word mover's distance. *Advances in neural information processing systems*, 29.

Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

Kingma, D. P.; and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Krizhevsky, A.; and Hinton, G. 2009. Learning multiple layers of features from tiny images. *Handbook of Systemic Autoimmune Diseases*, 1(4).

Kusner, M.; Sun, Y.; Kolkin, N.; and Weinberger, K. 2015. From word embeddings to document distances. In *International conference on machine learning*, 957–966. PMLR.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 2002. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278–2324.

Monge, G. 1781. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, 666–704.

Narwaria, M.; Lin, W.; McLoughlin, I. V.; Emmanuel, S.; and Chia, L.-T. 2012. Fourier transform-based scalable image quality measure. *IEEE Transactions on Image Processing*, 21(8): 3364–3377.

Peyré, G.; Cuturi, M.; et al. 2019. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6): 355–607.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Sisodia, D. S.; and Verma, S. 2011. Image pixel intensity and artificial neural network based method for pattern recognition. *World Acad. Sci. Eng. Technol*, 57: 742–745.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33: 7462–7473.

Solomon, J.; De Goes, F.; Peyré, G.; Cuturi, M.; Butscher, A.; Nguyen, A.; Du, T.; and Guibas, L. 2015. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4): 1–11.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Zeng, W.; and Gu, X. D. 2013. *Ricci flow for shape analysis and surface registration: theories, algorithms and applications*. Springer Science & Business Media.

Zeng, W.; Lui, L. M.; and Gu, X. 2014. Surface registration by optimization in constrained diffeomorphism space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4169–4176.

Zeng, W.; Lui, L. M.; Luo, F.; Chan, T. F.-C.; Yau, S.-T.; and Gu, D. X. 2012. Computing quasiconformal maps using an auxiliary metric and discrete curvature flow. *Numerische Mathematik*, 121(4): 671–703.

Zeng, W.; Luo, F.; Yau, S.-T.; and Gu, X. D. 2009. Surface quasi-conformal mapping by solving Beltrami equations. In *IMA International Conference on Mathematics of Surfaces*, 391–408. Springer.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhang, X.; Ge, X.; Xu, T.; He, D.; Wang, Y.; Qin, H.; Lu, G.; Geng, J.; and Zhang, J. 2024. Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. In *European Conference on Computer Vision*, 327–345. Springer.

Zhu, L.; Lin, G.; Chen, J.; Zhang, X.; Jin, Z.; Wang, Z.; and Yu, L. 2025. Large Images Are Gaussians: High-Quality Large Image Representation with Levels of 2D Gaussian Splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 10977–10985.