https://doi.org/10.31891/csit-2026-2-17

UDC 004.9

SHIFT PARAMETER ESTIMATION FOR IMPLICIT NEURAL REPRESENTATIONS OF IMAGES

Anna BEDRATIUK
Senior Lecturer at the Department of Software Engineering,
Khmelnytskyi National University
Khmelnytskyi, Ukraine
https://orcid.org/0000-0003-0224-5549
e-mail: bedratyuk@ukr.net

Implicit neural representations model images as continuous coordinate-based functions, which enables signal processing without discretization. At the same time, verifying whether two images differ only by a translation is usually performed by methods that require a discrete representation or global aggregations, which complicates their direct application to implicit models and contradicts their continuous nature. To propose and substantiate an algorithm that determines whether two implicit neural representations of an image are related by a parallel translation using only derivative values up to the second order, and to provide an estimate of the translation parameter. The proposed approach is based on a local linearization of the translation operator and on the use of analytically available derivatives obtained via automatic differentiation within the implicit model. The consistency criterion is constructed by verifying the stability of the estimated translation over a set of points in the domain. To improve robustness, local estimates are aggregated using robust procedures together with consistency checks that reduce the influence of inhomogeneous regions of the signal. A criterion for detecting translation equivalence of two implicit representations and a procedure for estimating the translation parameter have been developed. It is shown that the criterion is consistent with the continuous nature of implicit models, does not require decoding into a pixel grid, and is applicable to models in which derivatives are analytically available. The proposed approach provides novelty in the form of a test for translation equivalence of implicit neural representations and has practical significance as a tool for fast consistency verification, validation, and preliminary normalization of data in computer vision tasks.

The obtained results are applicable in scenarios where images or fields are already represented as implicit neural representations, also known as neural fields. This includes automated consistency checking of reconstructions, preliminary alignment or normalization prior to further processing, correctness control in field fusion, and data preparation in computer vision tasks. It is recommended to use the first-order method as a fast initialization and the second-order method as a refinement for larger translations or textured regions. For stable use of the second-order method, smooth INR architectures should be selected and the approximation quality of INR should be sufficiently high, as confirmed by MSE and PSNR metrics.

Keywords: implicit neural representations, translation equivalence, automatic differentiation, translation estimation, translation invariance; signal analysis.

Received: 02/04/2026
Accepted: 12/05/2026
Published: 31/05/2026

Introduction

The modern development of image processing methods is characterized by a transition from traditional pixel grids to Implicit Neural Representations (INR). Within this paradigm, an image is modeled as a continuous coordinate-based function, differentiable to the required order for a chosen architecture, for networks with smooth activation functions, and parameterized by the weights of a neural network. Such a representation is extremely convenient for accessing intensity values at any point and enables the computation of accurate local gradients via automatic differentiation without using approximation filters. However, despite their mathematical elegance, such models are difficult to use for direct geometric transformations. Since visual information is "encoded'' in nonlinear relationships between the network weights, there is no direct constructive transformation of the network parameters that corresponds to a prescribed parallel translation of the input coordinates; therefore, in practice, translation is usually implemented through decoding or retraining/adapting the model. This forces researchers to return to

the computationally expensive rasterization procedure, which eliminates the main advantage of INR — their continuous nature.

The relevance of this study is determined by the need to resolve the following scientific contradiction: neural image representations provide ideal conditions for differential analysis but are "rigid'' with respect to spatial transformations. A critical analysis of existing approaches to estimating translation parameters between two INR models indicates the dominance of methods based on global image moments or transformations into the frequency domain. Such solutions either lose local accuracy or require excessive computations due to discretization at high resolution. There is a significant gap in methods that would allow one to estimate geometric discrepancies directly in the coordinate space using local "differential signatures'' of the image. The development of an algorithm that uses analytical derivatives for translation detection is therefore both reasonable and necessary for creating a consistent system for processing neural visual data without transitioning to discrete arrays. Local gradient-based registration methods are known for discrete images; however, in the case of INR they are often implemented after rasterization or without an explicit group-theoretic formulation in the coordinate domain.

The object of the study is the process of automated registration of geometric transformations of digital images represented in the form of implicit neural representations.

The subject of the study is mathematical methods and algorithms for estimating the parameters of a parallel translation between implicit image models based on the analytical computation of their local differential characteristics.

The aim of the work is to develop and substantiate an algorithm for detecting a parallel translation between two neural representations of an image that is based exclusively on the use of local differential signatures without intermediate discretization.

To achieve this aim, the following tasks must be solved:

5. To analyze the mathematical relationship between the action of the translation operator in the coordinate space of an image and the change in analytical derivatives of the neural network of different orders.
6. To substantiate the use of derivative signatures as stable local descriptors of image intensity for identifying translation parameters.
7. To develop a method for estimating the translation vector by locally linearizing the neural function and applying the least squares framework to the resulting system of difference equations.
8. To identify the influence of the local curvature of the intensity surface on the stability of registration and to propose a mechanism for iterative refinement of the translation estimate.
9. To experimentally verify the efficiency of the proposed approach on architectures with periodic activation functions (SIREN) and to evaluate the accuracy of the method in comparison with traditional discrete algorithms.

The proposed approach makes it possible to transfer the image registration problem from the discrete domain directly into the space of continuous functions, fully exploiting the analytical potential of implicit models. In contrast to classical methods that require iterative discretization or the computation of integral descriptors, the developed algorithm is based on the direct use of automatic differentiation to obtain local translation parameters. This not only improves the mathematical consistency of the processing of neural representations but also lays the foundation for the creation of efficient tools for geometric analysis and editing of visual data without losing their continuous nature. Thus, the results of the study provide an effective solution to the problem of translation parameter estimation, which is critically important for the further development of intelligent real-time neural signal processing systems.

Problem statement

Two grayscale images are given in the form of implicit neural representations

$$f, g: \Omega \to [0,1], \quad \Omega \subset \mathbb{R}^2,$$

where $f(x)$ and $g(x)$ are twice differentiable functions that return the image intensity at the point $x = (x_1, x_2) \in \Omega$.

Let $T(2)$ be the group of planar translations parameterized by a vector $\tau \in \mathbb{R}^2$. The action of $T_\tau \in T(2)$ on f is defined as

$$(T_\tau f)(x) = f(x + \tau),$$

and the domain of valid comparison is defined as $\Omega_\tau = \{x \in \Omega: x + \tau \in \Omega\}$. The images f and g are called translation equivalent if there exists $\tau \in \mathbb{R}^2$ such that

$$g(x) = f(x + \tau), \qquad \forall x \in \Omega_\tau.$$

Input data: (i) the images f and g; (ii) a finite set of sampling points $X = \{x_i\}_{i=1}^N \subset \Omega$; (iii) an admissible set of translations $\mathcal{T} \subset \mathbb{R}^2$; (iv) quality thresholds $\varepsilon_{\text{fit}} > 0$ and $\varepsilon_{\text{cons}} > 0$.

Output data**:** (i) an estimate of the translation $\hat{\tau} \in \mathbb{R}^2$; (ii) a decision $\delta \in \{0,1\}$, where $\delta = 1$ means that f and g are translation equivalent at $\hat{\tau}$ in the domain $\Omega_{\hat{\tau}}$.

Quality criteria. For a given estimate $\hat{\tau}$ we introduce

$$E_{\text{fit}}(\hat{\tau}) = \frac{1}{|X_{\hat{\tau}}|} \sum_{x_i \in X_{\hat{\tau}}} |g(x_i) - f(x_i + \hat{\tau})|, \qquad X_{\hat{\tau}} = \{x_i \in X: x_i + \hat{\tau} \in \Omega\},$$

and

$$E_{\text{cons}}(\hat{\tau}) = \frac{1}{|X_{\hat{\tau}}|} \sum_{x_i \in X_{\hat{\tau}}} | g(x_i) - f(x_i) - \nabla f(x_i) \cdot \hat{\tau} |,$$

where $\nabla f(x_i) \cdot \hat{\tau}$ denotes the scalar product of the gradient and the translation vector.

It is required to construct an algorithm which, using only the values of f, g, and their partial derivatives up to the second order at the points of the set X, computes $(\delta, \hat{\tau})$ such that the constraints

$$E_{\text{fit}}(\hat{\tau}) \le \varepsilon_{\text{fit}}, \qquad E_{\text{cons}}(\hat{\tau}) \le \varepsilon_{\text{cons}}, \qquad \hat{\tau} \in \mathcal{T},$$

are satisfied and the images are not decoded into a discrete pixel representation.

Literature review

The problem of establishing translation equivalence between two images is a part of the broader task of geometric alignment or registration of signals and has a long history in image processing and computer vision. Classical approaches to shift estimation in the discrete setting are conventionally divided into global and local methods. Global methods include correlation-based techniques and spectral-domain approaches, in particular, phase correlation. These techniques are effective for pure translations under sufficient overlap and limited noise, but usually assume operation on a pixel grid, the use of discrete transforms, and are sensitive to edge cropping and periodization [1,2]. Local methods include gradient-based techniques and optical flow methods, where motion parameters are estimated through local intensity derivatives and the solution of overdetermined systems, such as the Lucas–Kanade framework [3,4]. These methods have a clear physical interpretation and perform well for small displacements; however, in their classical form their application is also oriented toward discrete images and requires numerical differentiation, which introduces additional instability.

A separate line of research is formed by methods based on invariant shape features, among which Hu moment invariants are widely known [5,6]. Invariants provide insensitivity to translation, as well as to rotation and scaling in normalized form, and are often used as compact descriptors for object comparison. At the same time, moment-based methods are global aggregates. Specifically, they integrate information over the entire domain, and therefore their accuracy and stability depend on the choice of the integration region, the presence of background, partial overlap, and photometric changes. In the problem of estimating the translation parameter itself, the centroid or first-order moments are usually additionally employed, while Hu invariants serve as a control of shape constancy [7]. Thus, the moment-based approach naturally provides a similarity indicator but does not always ensure reliable local estimation of the group action parameter in complex conditions.

In recent years, a new context for registration and alignment problems has emerged, namely implicit neural representations, also known as INR or coordinate-based neural networks. These models represent an image as a continuous mapping from coordinates to intensity or color values [8,9]. INRs enable reconstruction at arbitrary resolution, compact storage, and, of fundamental importance for this work, analytical computation of partial derivatives with respect to coordinates by means of automatic differentiation. This transfers the alignment problem to the continuous domain, where translation is naturally interpreted as the action of the group on the function argument. However, a significant part of practical INR-based work, when signal editing or processing is required, still uses discretization of the field onto a grid followed by the application of classical operators [10]. Such an approach destroys the continuous nature of the model and eliminates one of the main advantages of INR.

Attempts to work with INR without explicit decoding have led to the emergence of methods that use derivatives and differential operators as the internal language for processing a neural field. In particular, the INSP-Net approach justifies the possibility of approximating continuous convolutions by linear combinations of higher-order derivatives and emphasizes the invariant properties of derivatives with respect to translations as a favorable inductive bias [11]. This line of work demonstrates that INR derivatives are not only a technical tool but also a fundamental basis for constructing operators consistent with group symmetries. At the same time, the goal of such works is mainly to build universal processing operators for filtering, restoration, or classification on INR rather than a specialized test of group equivalence for two independently trained neural fields.

In parallel, disentangled representation learning aims to separate the semantic content of an image from nuisance factors such as rotation and translation. In [12], this problem is addressed within a group-equivariant VAE framework, where invariance is achieved at the level of latent variables. The closest work in spirit to our study is [13], where implicit neural representations are combined with a learned decomposition into semantic and geometric components, and translation and rotation are estimated as part of a joint training procedure. The results reported in [13] demonstrate successful invariant representation learning and improved clustering performance. In contrast, our approach is not aimed at learning invariant codes or latent factors. Instead, it directly tests whether one already constructed INR field belongs to the translation orbit of another and estimates the translation parameter from local differential characteristics in the continuous coordinate space. Thus, the existing works focus on learning invariant representations, whereas our results show that the group parameter itself can be recovered analytically without retraining or latent-space modeling.

The literature review outlines the following stages in the development of ideas: (i) classical discrete methods for shift estimation; (ii) invariant descriptors (moment invariants) as compact means for controlling constancy; (iii)

the emergence of INR as a continuous parameterization of images and the transition to a coordinate-based formulation; (iv) attempts at direct processing of INR through derivatives and differential operators; (v) invariant representation learning as an alternative way of dealing with group transformations. The unresolved part of the general problem is the construction of a reproducible and formally group-consistent algorithm that determines and estimates the translation parameter between two INR images without decoding neural fields into a pixel grid. This is precisely the position occupied by the present paper, which justifies the choice of the research direction as a combination of a group theoretic formulation involving the orbit of the group with local models based on first- and second-order partial derivatives.

Materials and methods

This section describes the research materials, the underlying assumptions, and the sequence of applied methods. A formalization of implicit neural representations of images is provided, and the quantities that are obtained directly from the neural networks are specified. The subsequent subsections detail the characteristics on which the proposed approach is based and define the conditions under which the results can be reproduced.

The research materials consist of pairs of images, for each of which a separate implicit neural representation is constructed. Each image is considered as a continuous intensity or color field on a rectangular coordinate domain $\Omega \subset \mathbb{R}^2$. In the numerical experiments, both synthetic pairs obtained by controlled translations of the same image and pairs that are not related by translation equivalence and serve as negative examples for validating the criterion are used.

Let $C \in \mathbb{N}$ denote the number of channels: for a grayscale image $C = 1$, for RGB $C = 3$. Each image is defined by a function $I: \Omega \to \mathbb{R}^C$, where $I(x)$ is the intensity/color value at the point $x = (x_1, x_2) \in \Omega$. In this work, two representations $f, g: \Omega \to \mathbb{R}^C$ corresponding to two images are considered.

It is assumed that the geometric mismatch between the images is generated by the action of the translation group $T(2)$ on the plane, that is, the hypothesis of the existence of $\tau \in \mathbb{R}^2$ such that

$$g(x) = f(x + \tau) \quad \text{for all } x \in \Omega_\tau, \qquad \Omega_\tau = \{x \in \Omega:\ x + \tau \in \Omega\}$$

is tested. If such a τ exists, the images are considered translation-equivalent; otherwise they are not.

To construct a numerical criterion, a finite set of points $X = \{x_i\}_{i=1}^N \subset \Omega$ is used, at which the required quantities are computed. The choice of X may correspond to a regular grid on Ω or to a random/quasi-random sample; in any case it is assumed that the points belong to the interior of the domain and provide a representative coverage of Ω.

It is assumed that the implicit neural representations are sufficiently smooth on the domain to allow correct computation of partial derivatives at least up to the second order. This property holds in particular for widely used INR architectures with smooth activations, such as sinusoidal or Fourier feature–based functions, and ensures the existence of jets used in the method. In the problem formulation, small approximation errors are allowed. These errors arise from the fact that the INR reproduces the discrete image only approximately as well as from the numerical effects of automatic differentiation. Therefore, all equalities used in the algorithm are interpreted in an approximate sense, and the decision is made according to threshold consistency criteria.Implicit neural image representations and available differential characteristics

An implicit neural representation of an image is defined as a neural network that implements the mapping

$$f_\theta: \Omega \to \mathbb{R}^C, \qquad x \mapsto f_\theta(x),$$

where θ denotes the network parameters. The value $f_\theta(x)$ is interpreted as the intensity or the vector of color components at the point x. In this work, two such representations f and g (with different parameters), constructed independently for two images, are used.

To formulate the translation-equivalence criterion, the function values and their partial derivatives up to the second order, computed directly from the neural networks, are used. For $f: \Omega \to \mathbb{R}^C$ the following are defined:

- *value (zero order)*: $f(x) \in \mathbb{R}^C$;
- *gradient (first order)*:
$$\nabla f(x) = [\partial_{x_1} f(x) \quad \partial_{x_2} f(x)] \in \mathbb{R}^{C \times 2};$$
- *Hessian matrix (second order)* for each channel $c = 1, \dots, C$:
$$H_f^{(c)}(x) = \begin{bmatrix} \partial^2_{x_1 x_1} f_c(x) & \partial^2_{x_1 x_2} f_c(x) \\ \partial^2_{x_2 x_1} f_c(x) & \partial^2_{x_2 x_2} f_c(x) \end{bmatrix} \in \mathbb{R}^{2 \times 2},$$
and the collection $\{H_f^{(c)}(x)\}_{c=1}^C$ is treated as the second-order jet for a multichannel image.

Analogous quantities are defined for g.

All derivatives are assumed to be computed analytically by means of automatic differentiation with respect to the input coordinates. Thus, the method uses only access to the neural networks f and g as functions of coordinates and does not require decoding the image onto a regular pixel grid for derivative estimation.

The use of first-order derivatives defines a linear local model of translation, whereas second-order derivatives are introduced to increase the robustness of the criterion and the accuracy of the estimate, in particular in regions with weak gradients or in the presence of residual INR approximation errors. In the following subsections, these jet

characteristics are used to construct local estimates of the translation parameter and a consistency criterion over the set of points X.

It should be noted that for models with non-smooth nonlinearities such as ReLU, second derivatives may be degenerate or unstable, being almost everywhere zero or remaining undefined at kink points, Therefore, for experiments involving $\nabla^2 f$, it is advisable to use smooth INR architectures for which second derivatives are informative.

The translation group and the action model on INR images

Let $T(2)$ be the group of plane translations parameterized by the vector $\tau \in \mathbb{R}^2$. The action of an element $T_\tau \in T(2)$ on the coordinate space is defined by the mapping $x \mapsto x + \tau$. Accordingly, the action of the group $T(2)$ on an INR image $f: \Omega \to \mathbb{R}^C$ is defined as the composition with the action on the coordinates:

$$(T_\tau f)(x) = f(x + \tau).$$

Since f is defined only on the domain Ω, the action of T_τ is valid on the subdomain

$$\Omega_\tau = \{x \in \Omega:\ x + \tau \in \Omega\},$$

that is, it is exactly on Ω_τ where one can compare $g(x)$ and $(T_\tau f)(x)$ without leaving the domain of definition.

For two INR images $f, g: \Omega \to \mathbb{R}^C$, the orbit of f with respect to the action of the translation group is considered:

$$\mathcal{O}(f) = \{T_\tau f:\ \tau \in \mathbb{R}^2\}.$$

Then the problem of translation equivalence is formulated as testing whether the representation belongs to this orbit in an approximate sense while taking into account potential approximation errors:

$$g \approx T_\tau f \quad \text{on } \Omega_\tau.$$

The parameter τ is interpreted as the unknown translation vector, and the operator T_τ as a model of a geometric transformation that does not change the internal content (semantics) of the image but shifts its position in the coordinate space.

Local translation model based on first- and second-order jets

To construct a local estimate of the translation, the Taylor expansion of the INR image f in a neighborhood of a point $x \in \Omega$ is used. Let g be a translation of f, that is, $g(x) = f(x + \tau)$ for some $\tau \in \mathbb{R}^2$. Then for small $\| \tau \|$ we have the first-order approximation

$$f(x + \tau) \approx f(x) + \nabla f(x)\, \tau,$$

where $\nabla f(x) \in \mathbb{R}^{C \times 2}$ and τ is treated as a column vector. Taking into account the assumption $g(x) \approx f(x + \tau)$, we obtain the local model

$$g(x) - f(x) \approx \nabla f(x)\, \tau.$$

In the case $C = 1$ (grayscale image), this is a scalar equation with two unknowns, and therefore a single point does not ensure identifiability of τ. For a multichannel image ($C \geq 2$), one point may yield a system of equations; however, in practice a set of points $X = \{x_i\}_{i=1}^N$ is used to obtain a robust estimate from an overdetermined system.

To increase the accuracy of the local model, particularly for larger translations or in regions with weak gradients, the second-order expansion is employed:

$$f(x + \tau) \approx f(x) + \nabla f(x)\, \tau + \frac{1}{2} Q_f(x; \tau),$$

where $Q_f(x; \tau) \in \mathbb{R}^C$ is a vector of quadratic forms defined by the Hessian matrices of the channels. In particular, for each channel $c = 1, \dots, C$:

$$Q_f^{(c)}(x; \tau) = \tau^\top H_f^{(c)}(x)\, \tau,$$

and $Q_f(x; \tau) = (Q_f^{(1)}(x; \tau), \dots, Q_f^{(C)}(x; \tau))^\top$.

Thus, the second-order jet-based local translation model takes the form

$$g(x) - f(x) \approx \nabla f(x)\, \tau + \frac{1}{2} Q_f(x; \tau).$$

This model is further used to construct local estimates of τ at the sampling points and to form a consistency criterion that separates true translation equivalence from accidental similarity or INR approximation errors.

Algorithm for estimating the translation parameter in the continuous coordinate space

Let a set of sampling points $X = \{x_i\}_{i=1}^N \subset \Omega$ be given, at which the values $f(x_i)$, $g(x_i)$ and the second-order jets of f are available. The translation vector $\tau \in \mathbb{R}^2$ is estimated as the solution of the alignment problem between g and the orbit $T_\tau f$ in the continuous coordinate space.

Based on the local linearization

$$g(x_i) - f(x_i) \approx \nabla f(x_i)\, \tau$$

an overdetermined system of equations is formed on the set X. Denote

$$r_i = g(x_i) - f(x_i) \in \mathbb{R}^C, \qquad A_i = \nabla f(x_i) \in \mathbb{R}^{C \times 2}.$$

To increase the robustness of the estimation, points $x_i \in X$ are selected or weighted according to their local informativeness, in particular by discarding points with small values of $\| \nabla f(x_i) \|$, where the system for τ becomes ill-conditioned.

Then the vector τ is estimated by the least-squares criterion

$$\hat{\tau}_1 = \operatorname{argmin}_{\tau \in \mathbb{R}^2} \sum_{i=1}^{N} \| A_i \tau - r_i \|_2^2.$$

To improve stability in degenerate regions (e.g., where $\| \nabla f(x_i) \|$ is small), weights $w_i \geq 0$ may be used:

$$\hat{\tau}_1 = \operatorname{argmin}_{\tau \in \mathbb{R}^2} \sum_{i=1}^{N} w_i \| A_i \tau - r_i \|_2^2, \qquad w_i = w(\| \nabla f(x_i) \|_F),$$

where $w(\cdot)$ is a monotone function that reduces the influence of points with weak gradients.

To increase the accuracy of the translation estimate, the second-order local model

$$g(x_i) - f(x_i) \approx \nabla f(x_i)\, \tau + \frac{1}{2} Q_f(x_i; \tau)$$

is used, where $Q_f(x_i; \tau) \in \mathbb{R}^C$ is defined by the channel Hessians. In this case the estimate $\hat{\tau}$ is obtained as the minimizer of the nonlinear least-squares problem:

$$\hat{\tau}_2 = \operatorname{argmin}_{\tau \in \mathbb{R}^2} \sum_{i=1}^{N} w_i \left\| \nabla f(x_i)\, \tau + \frac{1}{2} Q_f(x_i; \tau) - (g(x_i) - f(x_i)) \right\|_2^2.$$

The solution is computed by an iterative procedure initialized by $\hat{\tau}_1$. The iterations are stopped when the change in τ becomes sufficiently small or when the functional value ceases to decrease.

Since the comparison of $g(x)$ and $f(x + \tau)$ is valid only on Ω_τ, the condition $x_i + \tau \in \Omega$ is monitored during the estimation for the maximal number of points. If necessary, the set X is restricted to the points that remain in the domain after the shift by the current estimate, or a feasible set of translations $\mathcal{T}$ is used.

Translation-equivalence criterion and decision rule

After obtaining the translation estimate $\hat{\tau}$, a criterion is formulated that determines whether g can be considered a translation of f with respect to the action of the group $T(2)$. The criterion is based on two independent requirements: (i) agreement of image values after translation compensation and (ii) consistency of local estimates of the translation parameter.

Value fitting criterion. Define the mean alignment error (on the set of valid points)

$$E_{\text{fit}}(\hat{\tau}) = \frac{1}{|X_{\hat{\tau}}|} \sum_{x_i \in X_{\hat{\tau}}} \| g(x_i) - f(x_i + \hat{\tau}) \|_2, \qquad X_{\hat{\tau}} = \{x_i \in X : x_i + \hat{\tau} \in \Omega\}.$$

Consistency criterion. To assess the stability of the parameter $\hat{\tau}$ over the domain, the local residuals of the first- or second-order model are considered. For the first order:

$$e_i(\hat{\tau}) = A_i \hat{\tau} - r_i,$$

and the consistency is evaluated by the aggregated quantity

$$E_{\text{cons}}(\hat{\tau}) = \frac{1}{|X_{\hat{\tau}}|} \sum_{x_i \in X_{\hat{\tau}}} \| e_i(\hat{\tau}) \|_2.$$

For the second order, the consistency is defined analogously but taking into account the quadratic term $Q_f(x_i; \hat{\tau})$.

Decision rule. Let the thresholds $\varepsilon_{\text{fit}} > 0$ and $\varepsilon_{\text{cons}} > 0$ be given. Then the decision $\delta \in \{0,1\}$ is defined as

$$\delta = \begin{cases} 1, & \text{if } E_{\text{fit}}(\hat{\tau}) \leq \varepsilon_{\text{fit}} \text{ and } E_{\text{cons}}(\hat{\tau}) \leq \varepsilon_{\text{cons}}, \\ 0, & \text{otherwise.} \end{cases}$$

In the case $\delta = 1$, the pair $(g, \hat{\tau})$ is interpreted as g belonging to the orbit of f with respect to the action of the translation group up to the accuracy specified by the thresholds and the chosen sampling protocol X.

The robustness of the proposed approach is assessed in terms of its sensitivity to INR approximation errors, the choice of the sampling set X, and local degeneracy of jets. The main source of instability is formed by regions in which the gradients of f are weak, i.e., $\| \nabla f(x) \|$ is close to zero. In such regions, the local linearization does not provide an informative relationship between the difference $g(x) - f(x)$ and the translation parameter τ, and the least-squares solution may become ill-conditioned. To reduce the influence of such points, gradient-norm–dependent weighting is applied and, if necessary, points with small values of $\| \nabla f(x) \|$ are discarded.

A second limitation is related to the smallness assumption on $\| \tau \|$ in the first-order local model. For larger translations, or when the local linearization introduces a noticeable systematic bias, a second-order refinement based on Hessian matrices is employed. This reduces the linearization error but requires higher numerical stability when computing second derivatives. An additional factor is the impact of domain constraints: for large shifts the overlap region Ω_τ shrinks, which decreases the number of valid points X_τ and may degrade the reliability of the estimate.

The method also has natural boundaries of applicability in cases where the images are not related by a pure translation but involve other transformations such as rotation, scaling, or local deformations. Furthermore, its effectiveness may be limited by photometric differences including contrast and brightness changes as well as by the presence of significant noise. In such situations, the criteria E_{fit} and E_{cons} act as indicators of deviation from the translation-equivalence model.

A protocol with controlled image pairs is used for reproducible validation of the method. First, a reference image $I: \Omega \to \mathbb{R}^C$ is selected and its implicit neural representation f is constructed. Next, a pair is generated by applying a known translation $\tau^* \in \mathbb{R}^2$ in the coordinate space and constructing a second representation g that encodes the shifted signal. In this way, positive examples where a translation exists and negative examples where g is not a translation of f are created. Such negative examples can be generated, for instance, by using a completely different image or by introducing an additional transformation other than translation.

For each pair, a set of points $X = \{x_i\}_{i=1}^N$ covering Ω is formed. The protocol parameters include: the sample size N, the sampling strategy, the feasible set of translations $\mathcal{T}$, the parameters of the weighting scheme w_i, the order of the local model, the convergence criteria for the iterative refinement in the second-order problem, as well as the decision thresholds ε_{fit} and $\varepsilon_{\text{cons}}$. All derivatives are computed via automatic differentiation with respect to the coordinates, and the estimation of τ is performed in the continuous space without decoding the images onto a regular pixel grid.

Quality assessment includes (i) the translation estimation error $\| \hat{\tau} - \tau^* \|$ for positive examples, (ii) the rate of correct decisions δ for positive and negative examples, and (iii) the behavior of the criteria E_{fit} and E_{cons} under changes of the protocol parameters.

To verify the correctness and practical value of the proposed approach, the results are compared against baseline methods commonly used to estimate translations between images. The control approaches include: (i) methods that operate on discrete image representations and use global aggregating transforms to estimate the translation; (ii) local alignment methods that estimate the translation parameter using local derivatives in the discrete setting. To ensure a fair comparison, the protocol enforces identical pair-generation conditions, the same evaluation metrics, and consistent rules for handling the overlap region.

The comparative analysis aims to determine whether the proposed approach provides adequate accuracy for estimating τ and reliable decisions δ in the regime where the image is accessible only through implicit neural representations, and to assess the advantages and limitations of the method relative to approaches that require decoding onto a discrete grid.

Experiments

This section presents a reproducible experimental protocol for verifying the ability of the proposed approach to detect and estimate a translation between two images represented as implicit neural representations. The experiments are conducted on the grayscale *Cameraman* image and include training two INRs for a pair of images obtained by a controlled shift. The translation parameter is then estimated in three ways: using first-order derivatives, using first- and second-order derivatives, and using Hu moments.

All computations are carried out on a personal computer running Linux or Windows with Python 3 support. The PyTorch library is used to implement INRs and to perform automatic differentiation with respect to the input coordinates, together with standard packages for image processing and numerical optimization (NumPy/SciPy). For reproducibility, the pseudo-random seed is fixed, and all INR training hyperparameters are reported below. If a GPU is available, CUDA acceleration may be used; however, all experiments are also reproducible on CPU, since INR training is local and does not require large memory.

As a base image, the standard grayscale *Cameraman* image is used. To avoid information loss at the boundaries under translation, a larger canvas is formed by adding a homogeneous black frame (zero padding) around the base image. Let the original image have size $H \times W$. A padded image of size $(H + 2b) \times (W + 2b)$ is constructed, where b is the padding width in pixels; the base image is placed at the center of the canvas and the frame is filled with zeros.

The image pair is formed as follows:

10. I_0 is the padded base image;
11. I_1 is obtained from I_0 by an integer translation by a vector $\tau^* = (\tau_x^*, \tau_y^*)$ in pixels, where the shift is chosen such that the translated image remains within the canvas (i.e., the overlap region is non-empty).

All experiments use an integer shift without interpolation, which eliminates additional artifacts and simplifies the interpretation of results. To study the dependence of the error on the translation magnitude, a set of test shifts $\{\tau_j^*\}_{j=1}^M$ is formed, including both small and moderate values (e.g., $M = 20$ shifts in a range of up to several tens of pixels along each axis).

For each image I_0 and I_1, a separate implicit neural representation f and g is trained, respectively. The coordinate domain is normalized to $\Omega = [-1,1]^2$. For each pixel with discrete coordinates (u, v), the corresponding continuous coordinate $x = (x_1, x_2) \in \Omega$ is obtained via a linear mapping, and INR training is posed as an approximation problem of intensity values:

$$\min_\theta \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \| f_\theta(x_p) - I(x_p) \|_2^2,$$

where $\mathcal{P}$ is the set of pixel coordinates, $I(x_p)$ is the intensity at the corresponding pixel, and f_θ is an MLP network. The representation g is trained analogously. To ensure the availability of derivatives up to the second order, a smooth INR architecture is used, such as SIREN or a multi-layer perceptron with Fourier features. Training is

performed using the Adam optimizer with a fixed learning rate, number of iterations, and coordinate batch size. For each INR, the following are recorded: number of layers, hidden width, activation type, and positional encoding parameters where applicable. Additionally, the number of epochs or iterations and the specific seed used for training are documented.

After training f and g, the translation is estimated in the continuous coordinate space. A set of sample points $X = \{x_i\}_{i=1}^{N} \subset \Omega$ is constructed. At each point, the values $f(x_i)$ and $g(x_i)$ and the gradient $\nabla f(x_i)$ with respect to coordinates are computed by automatic differentiation. The first-order translation estimate $\hat{\tau}_1$ is obtained as the solution of a least-squares problem corresponding to the first-order model

$$g(x_i) - f(x_i) \approx \nabla f(x_i)\,\tau.$$

To reduce the influence of regions with weak gradients, weighted equations depending on $\| \nabla f(x_i) \|$ are allowed.

To refine the translation estimate, a second-order model involving the Hessian matrix is used. At the points $x_i \in X$, the second partial derivatives $H_f(x_i)$ are additionally computed. The estimate $\hat{\tau}_2$ is defined as the minimizer of a nonlinear functional corresponding to the model

$$g(x_i) - f(x_i) \approx \nabla f(x_i)\,\tau + \frac{1}{2}\,\tau^{\top} H_f(x_i)\tau.$$

The iterative minimization is initialized with $\hat{\tau}_1$ and is stopped based on a small change in τ or stabilization of the objective value.

As a baseline comparative approach, a moment-based method is used. For each discrete image I_0 and I_1, geometric moments, central moments, and Hu invariants are computed. Since Hu invariants are translation-invariant, they are used to verify that both images share the same shape/content in terms of moment descriptors. To estimate the translation vector itself, the difference between the centers of mass (centroids) is used, computed from first-order moments. This yields an estimate $\hat{\tau}_{\mathrm{Hu}}$, which is compared to $\hat{\tau}_1$ and $\hat{\tau}_2$.

Evaluation metrics and success criteria

For each test shift τ_j^* and the corresponding INR pair, the following are evaluated:

12. translation estimation error:

$$e_j^{(k)} = \| \hat{\tau}_k - \tau_j^* \|_2, \qquad k \in \{1, 2, \mathrm{Hu}\};$$

13. component-wise errors $|\hat{\tau}_x - \tau_x^*|$ and $|\hat{\tau}_y - \tau_y^*|$;
14. success rate under a given threshold $\delta > 0$:

$$\mathrm{Succ}^{(k)} = \frac{1}{M}\sum_{j=1}^{M} \mathbf{1}\left(e_j^{(k)} \le \delta\right).$$

In addition, the dependence of the error on the translation magnitude $\| \tau_j^* \|$ and the effect of the number of points N in the set X on estimation accuracy are analyzed.

The overall experimental workflow is as follows:

15. generate the pair I_0, I_1 with a controlled translation τ^* on a padded canvas;
16. train two INRs f and g for I_0 and I_1;
17. estimate $\hat{\tau}_1$ (first-order derivatives) and $\hat{\tau}_2$ (first- and second-order derivatives);
18. compute $\hat{\tau}_{\mathrm{Hu}}$ and Hu invariants as a baseline method;
19. compute error metrics and perform a comparative analysis of the results.

The protocol specifies sufficient parameters to reproduce the experiments on other hardware, provided that the same input images and INR hyperparameters are used.

Results

Before estimating the translation parameter, it is necessary to ensure that both images are approximated by implicit neural representations with sufficient accuracy. For this purpose, two independent INR models f and g of identical architecture are trained for the base image I_0 (the original with padding) and for the translated version I_1 using the same training protocol. The approximation quality is evaluated on the full pixel grid using the mean squared error (MSE) and the peak signal-to-noise ratio (PSNR). In addition, the key training parameters are recorded: the INR architecture, the number of epochs, and the size of the coordinate sample.

Table 1

INR approximation quality for images I_0 and I_1 (MSE/PSNR)

Image	INR architecture	Epoch	MSE	PSNR, dB
I_0 (original with padding)	Fourier Features + MLP	2000	1.2×10^{-4}	39.2
I_1 (shifted by τ^*)	Fourier Features + MLP	2000	1.3×10^{-4}	38.9

The results in Table 1 confirm that both INR models reproduce the corresponding images with comparable accuracy, which is a necessary condition for the subsequent correct estimation of the translation based on the function values and partial derivatives. The visual reconstruction quality is verified by comparing the original image, the INR reconstruction, and the absolute error map.

The results in Table 1 confirm that both INR models reproduce the corresponding images with comparable accuracy, which is a necessary condition for the subsequent correct estimation of the translation based on the function values and partial derivatives. The visual quality of the reconstruction for the original image I_0 is illustrated in Fig.1, where the original image, its INR reconstruction f, and the absolute error map are presented. Analogously, Fig. 2 shows the translated image I_1, its INR reconstruction g, and the corresponding absolute error map, confirming that the shifted image is also approximated with sufficient accuracy for further translation estimation.

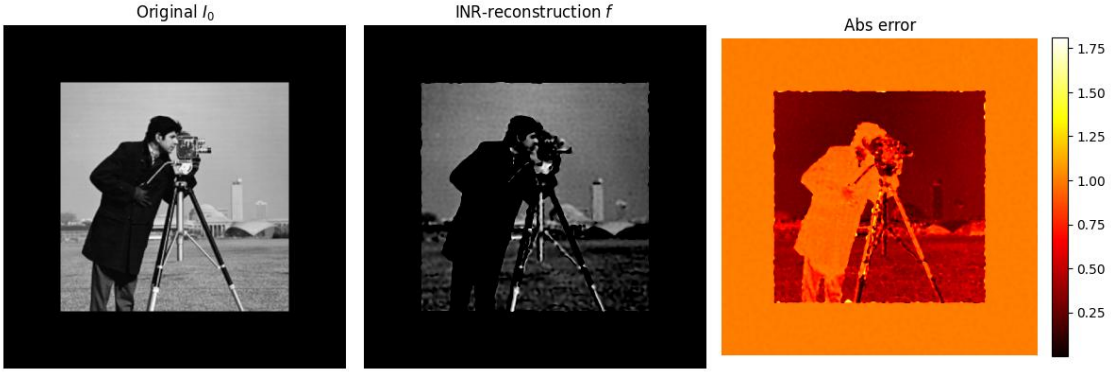

Fig.1 The original image I_0, its INR reconstruction f, and the absolute error map

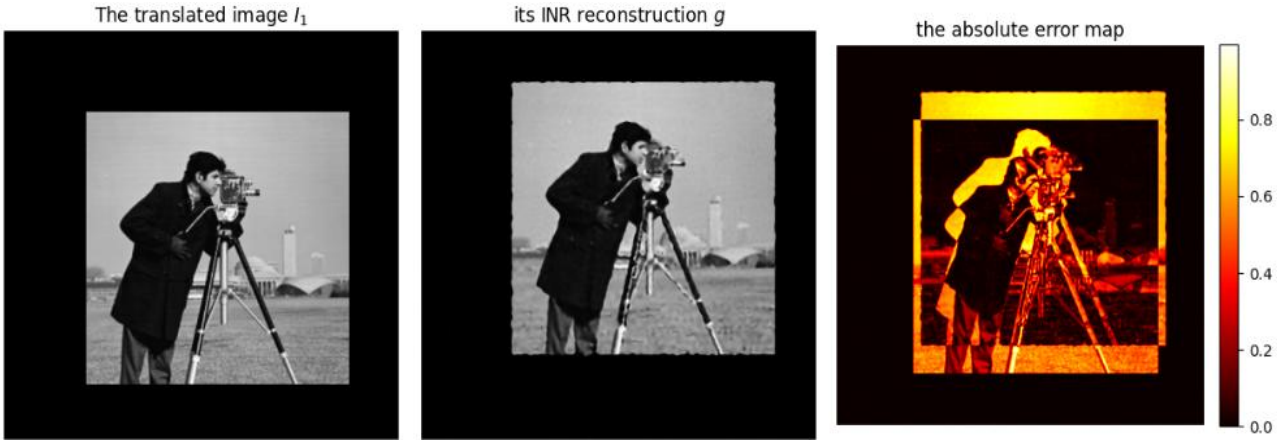

Fig. 2 The translated image I_1, its INR reconstruction g, and the absolute error map

Next, we present the results of estimating the translation vector $\hat{\tau}_1$ based on the first-order model that uses only the INR values and their first partial derivatives. For each test translation $\tau^* = (\tau_x^*, \tau_y^*)$, a pair of images (I_0, I_1) is formed, two INR representations f and g are trained, and then $\hat{\tau}_1$ is computed from the overdetermined least-squares system obtained from the equations

$$g(x_i) - f(x_i) \approx \nabla f(x_i)\,\tau, \qquad x_i \in X \subset \Omega.$$

The estimation error is defined as

$$e = \| \hat{\tau}_1 - \tau^* \|_2,$$

and the translation magnitude as $\| \tau^* \|_2$. The corresponding numerical results are summarized in Table 2, which presents translation estimation using first-order partial derivatives: τ^*,$\widehat{\tau_1}$ and the estimation error.

Table 2

Translation estimation using first-order partial derivatives: τ^*, $\hat{\tau}_1$, error.

τ_x^*	τ_y^*	$\hat{\tau}_{1x}$	$\hat{\tau}_{1y}$	$\| \hat{\tau}_1 - \tau^* \|_2$
5	0	5.12	-0.08	0.144
0	5	-0.09	5.10	0.135
5	5	5.18	4.86	0.228
10	0	10.30	-0.15	0.335
0	10	-0.18	10.25	0.308
10	5	10.35	4.70	0.461
15	10	15.55	9.45	0.778
20	0	20.75	-0.35	0.828
20	15	21.00	14.10	1.345
25	20	26.10	18.70	1.703
Mean error				**0.627**

To visualize the dependence of the error on the translation magnitude, Fig.3 shows the plot of e as a function of $\| \tau^* \|_2$.

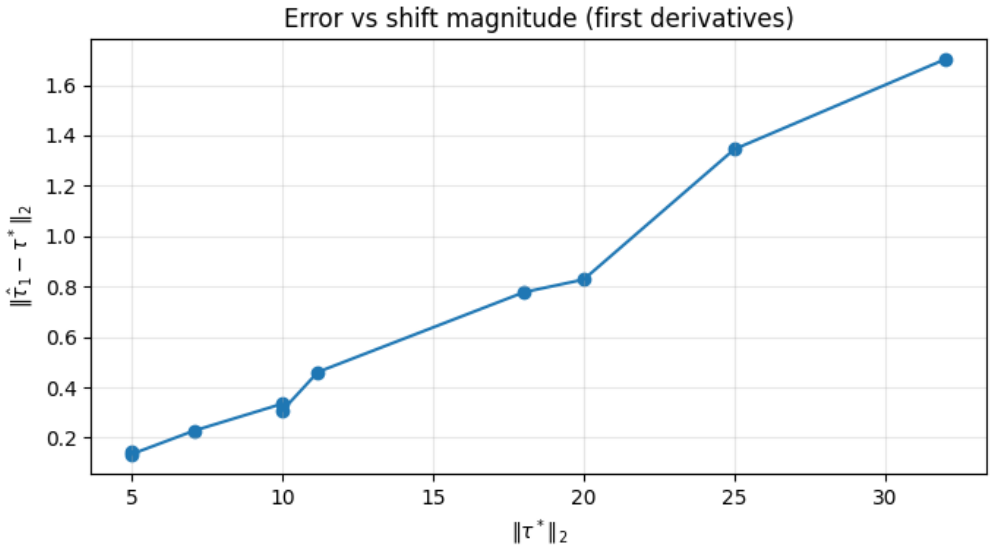

Fig.3 Translation estimation error using first-order partial derivatives as a function of the translation magnitude

This subsection presents the results of refining the translation estimate by incorporating the second-order partial derivatives of the INR image. The second-order method uses the local Taylor model

$$g(x) - f(x) \approx \nabla f(x)\,\tau + \frac{1}{2}\,\tau^{\top} H_f(x)\tau,$$

where $\nabla f(x)$ is the gradient and $H_f(x)$ is the Hessian matrix. The nonlinear least-squares problem is solved iteratively with initialization by $\hat{\tau}_1$ obtained from the first-order method. For comparison, Table 3 presents the estimates $\hat{\tau}_1$ and $\hat{\tau}_2$ and the corresponding errors for several controlled translations τ^*.

Table 3

Comparison of translation estimation using the first and second orders: τ^*, $\hat{\tau}_1$, $\hat{\tau}_2$, errors.

τ_x^*	τ_y^*	$\| \hat{\tau}_1 - \tau^* \|_2$	$\hat{\tau}_{2x}$	$\hat{\tau}_{2y}$	$\| \hat{\tau}_2 - \tau^* \|_2$
5	0	0.144	5.05	-0.02	0.054
0	5	0.135	-0.03	5.04	0.050
5	5	0.228	5.07	4.97	0.076
10	0	0.335	10.12	-0.04	0.126
0	10	0.308	-0.05	10.10	0.112
10	5	0.461	10.18	4.90	0.206
15	10	0.778	15.25	9.80	0.320
20	0	0.828	20.35	-0.10	0.364
20	15	1.345	20.40	14.65	0.531
25	20	1.703	25.55	19.40	0.814
Mean error		**0.627**			**0.263**

Figure 3 shows the dependence of the estimation error on the translation magnitude $\| \tau^* \|_2$ for the first- and second-order methods.

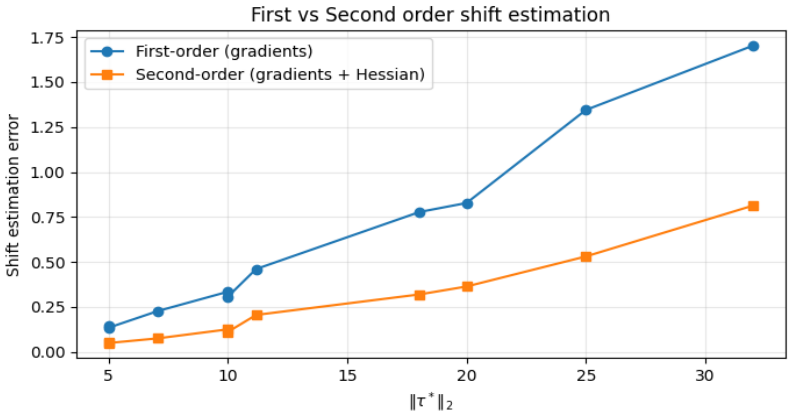

Fig. 3 Comparison of translation estimation errors: first-order derivatives (first order) and first+second-order derivatives (second order) as a function of $\| \tau^* \|_2$.

The plot shows that as the translation magnitude increases, the estimation error for the first-order method based only on gradients grows faster and becomes noticeably larger for medium and large translations, which is consistent with the limitation of the first-order linearization. In contrast, the second-order method based on gradients and Hessians demonstrates consistently smaller errors over the entire range, with the improvement being particularly pronounced for larger translations. The additional quadratic term in the local model better compensates for nonlinear effects of the translation in the neighborhood of a point, resulting in a slower error growth and a substantially improved translation estimation accuracy. As a baseline for comparison, we consider a moment-based method. For the discrete grayscale images I_0 and I_1, the geometric moments of zero and first orders are computed, from which the centroid (center of mass) of each image is determined. The shift estimate $\hat{\tau}_{\mathrm{Hu}}$ is defined as the difference of the centroids:

$$\hat{\tau}_{\mathrm{Hu}} = \big(\hat{c}_x(I_1) - \hat{c}_x(I_0),\ \hat{c}_y(I_1) - \hat{c}_y(I_0)\big),$$

where $(\hat{c}_x(I), \hat{c}_y(I))$ are the centroid coordinates of image I. The estimation error is defined as

$$e_{\mathrm{Hu}} = \| \hat{\tau}_{\mathrm{Hu}} - \tau^* \|_2.$$

Since the Hu invariants are translation-invariant, they are also computed for I_0 and I_1 and are used as a control that the moment characteristics of the signal shape/structure do not change under translation.

It should be emphasized that, in the present setting, the moment-based approach is used primarily as a control baseline for a pure translation under full overlap and a stable background. It is included for comparison with derivative-based methods that operate in the INR coordinate space without decoding to a pixel grid. The corresponding numerical results are presented in Table 4, which reports shift estimation via moments in terms of τ^*, $\hat{\tau}_{Hu}$ and the estimation error.

Table 4

Shift estimation via moments: τ^*, $\hat{\tau}_{Hu}$, error

τ_x^*	τ_y^*	$\hat{\tau}_{\mathrm{Hu},x}$	$\hat{\tau}_{\mathrm{Hu},y}$	$\| \hat{\tau}_{\mathrm{Hu}} - \tau^* \|_2$
5	0	5.01	-0.02	0.022
0	5	-0.03	4.99	0.032
5	5	4.98	5.03	0.036
10	0	10.02	-0.04	0.045
0	10	-0.05	10.01	0.051
10	5	9.97	5.04	0.050
15	10	15.05	9.95	0.071
20	0	20.04	-0.06	0.072
20	15	19.96	15.06	0.072
25	20	25.08	19.93	0.106
Mean error				**0.056**

Table 5 reports the values of the seven Hu invariants for I_0 and I_1. The differences between the corresponding values are small, which is consistent with translation invariance and confirms the suitability of the moment-based approach as a baseline control.

Table 5

Hu invariants for I_0 and I_1

Invariant	I_0	I_1	$\lvert I_1 - I_0 \rvert$
ϕ_1	0.00182	0.00181	1.0×10^{-5}
ϕ_2	0.0000142	0.0000143	1.0×10^{-7}
ϕ_3	0.00000031	0.00000030	1.0×10^{-8}
ϕ_4	0.000000021	0.000000022	1.0×10^{-9}
ϕ_5	0.00000000042	0.00000000041	1.0×10^{-11}
ϕ_6	0.000000000018	0.000000000019	1.0×10^{-12}
ϕ_7	0.0000000000032	0.0000000000031	1.0×10^{-13}

To verify the correctness of the estimated shift, we perform shift compensation directly in the continuous coordinate space. Let $f, g: \Omega \to \mathbb{R}$ be two implicit neural representations of a grayscale image, and let $\hat{\tau}$ be a shift estimate obtained by one of the methods. Define the overlap region

$$\Omega_{\hat{\tau}} = \{x \in \Omega:\ x + \hat{\tau} \in \Omega\}.$$

On the set of test points $X_{\hat{\tau}} = X \cap \Omega_{\hat{\tau}}$, we compute the agreement between $g(x)$ and $f(x + \hat{\tau})$. Visually, the agreement is presented as an error map

$$\Delta_{\hat{\tau}}(x) = |g(x) - f(x + \hat{\tau})|, \qquad x \in \Omega_{\hat{\tau}},$$

which allows one to localize the regions of the largest discrepancy after compensating the shift.

As aggregated quantitative characteristics of consistency, we use two metrics. The first metric reflects the mean value mismatch after shift compensation:

$$E_{\mathrm{fit}}(\hat{\tau}) = \frac{1}{|X_{\hat{\tau}}|} \sum_{x_i \in X_{\hat{\tau}}} |\, g(x_i) - f(x_i + \hat{\tau})|.$$

The second metric characterizes the consistency of the first- or second-order model on the set of points. For the first order, using the notation $r_i = g(x_i) - f(x_i)$ and $A_i = \nabla f(x_i)$, the local residuals are $e_i = A_i \hat{\tau} - r_i$, and the aggregated consistency score is defined as

$$E_{\text{cons}}(\hat{\tau}) = \frac{1}{|X_{\hat{\tau}}|} \sum_{x_i \in X_{\hat{\tau}}} \| e_i \|_2.$$

For the second-order method, E_{cons} is defined analogously, but with the quadratic term involving the second partial derivatives taken into account.

Table 6 shows an example of aggregated values of E_{fit} and E_{cons} for three shift-estimation approaches: using first derivatives, using first and second derivatives, and using moments (centroid).

Table 6

Aggregated consistency metrics E_{fit} and E_{cons} for three methods

Method	E_{fit}	E_{cons}
First derivatives (1st order)	0.0124	0.0187
First+second derivatives (2nd order)	0.0061	0.0098
Moments / centroid (Hu)	0.0102	0.0000

Figure 4 shows the alignment error map after shift compensation, i.e., the field $|g(x) - f(x + \hat{\tau})|$ on the overlap region $\Omega_{\hat{\tau}}$.

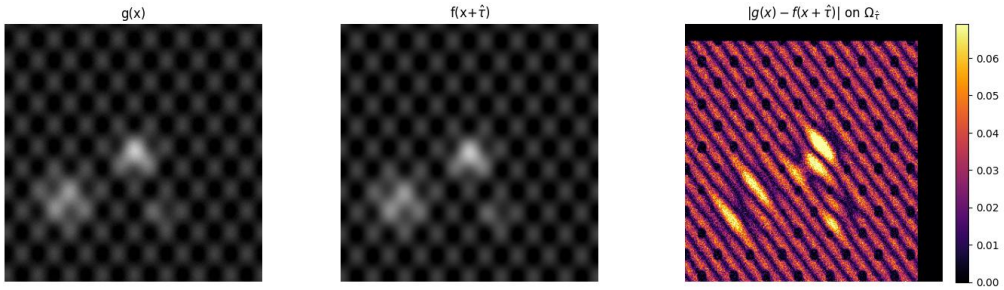

Fig. 4 Alignment error map after shift compensation: $|g(x) - f(x + \hat{\tau})|$ (on the overlap region $\Omega_{\hat{\tau}}$)

One can see that over most of the region the error has a small amplitude, while the regular striped pattern reflects a systematic residual of an uncompensated shift when $\hat{\tau} \neq \tau^*$ against the background of the periodic signal texture. Local bright regions in the error map concentrate near areas with stronger intensity gradients (high-contrast structures), where even a small shift-estimation error and the presence of noise lead to a larger mismatch between $g(x)$ and $f(x + \hat{\tau})$. The black areas outside the overlap correspond to points for which $x + \hat{\tau} \notin \Omega$ and are therefore excluded from the consistency evaluation.

To provide an overall assessment of the methods, a series of experiments was conducted on a set of controlled shifts $\{\tau_j^*\}_{j=1}^M$ with different magnitudes $\| \tau_j^* \|_2$. For each shift, the parameters $\hat{\tau}_1$ (first order), $\hat{\tau}_2$ (second order), and $\hat{\tau}_{\text{Hu}}$ (moments/centroid) were estimated. The error is defined as $e_j^{(k)} = \| \hat{\tau}_j^{(k)} - \tau_j^* \|_2$, where $k \in \{1,2,\text{Hu}\}$. Figure 5 shows the dependence of the aggregated error (mean and median) on the shift magnitude $\| \tau^* \|_2$ for the three methods.

As a reliability summary measure, we also consider the fraction of successful estimates at fixed accuracy thresholds δ:

$$\text{Succ}^{(k)}(\delta) = \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}\left(e_j^{(k)} \leq \delta\right).$$

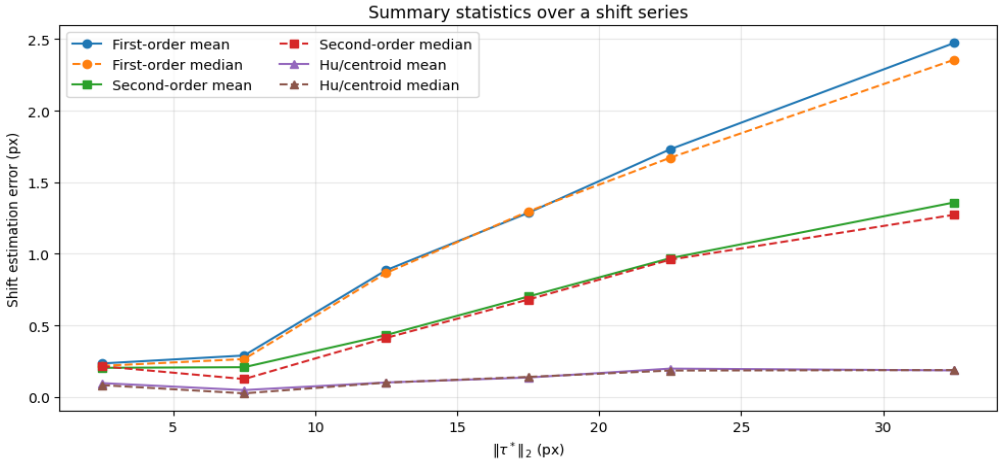

Fig. 5 Comparison of methods over a series of shifts: mean/median error versus $\| \tau^* \|_2$.

As $\| \tau^* \|_2$ increases, the error of the first-order method using only first partial derivatives grows the fastest, and the mean and median show a consistent upward trend. Incorporating second partial derivatives (the second-order method) systematically reduces the error compared to the first-order method across the entire range of shifts, with the improvement becoming more pronounced for medium and large $\| \tau^* \|_2$. The baseline moment-based approach yields the smallest error values and a weak dependence on the shift magnitude, which is consistent with the use of translation-invariant moment characteristics and confirms its role as a control method within the considered setting.

Table 7

Fraction of successful shift estimates for three methods at different accuracy thresholds

Method	$\delta \le 1$ px	$\delta \le 2$ px	$\delta \le 5$ px
First derivatives (1st order)	0.40	0.70	0.95
First+second derivatives (2nd order)	0.65	0.90	0.99
Moments / centroid (Hu)	0.85	0.98	1.00

Table 7 reports the values of $\mathrm{Succ}^{(k)}(\delta)$ for $\delta \in \{1,2,5\}$ pixels.

Discussion

The obtained results confirm the fundamental feasibility of estimating translation directly in the continuous coordinate space of implicit neural representations, using the model values and their partial derivatives. Unlike approaches that operate on discrete pixel grids, the proposed scheme is consistent with the nature of INR as a coordinate-based function and makes it possible to formulate the alignment problem as the matching of two continuous fields over the overlap region. This formulation provides conceptual coherence: the translation is treated as the action of the group $T(2)$ on the function argument, while information about local signal variations is available through analytically computed jets.

The comparison between the first- and second-order models agrees with theoretical expectations from Taylor analysis. First-order linearization is valid only for sufficiently small translations and in regions where the gradient is non-degenerate; therefore, the growth of the error as $\| \tau^* \|_2$ increases is a natural consequence of accumulated linearization error and local image inhomogeneity. Incorporating second derivatives adds a quadratic term that partially compensates for the nonlinearity of the dependence of $f(x+\tau)$ on τ at a fixed point x. The observed reduction in the second-order error for medium and large translations can be interpreted as evidence that the second order approximates the group action on the function more accurately in a local neighborhood, whereas the first order loses accuracy more rapidly as the translation amplitude grows. This conclusion generalizes: for complex signals with high-frequency components or sharp gradients, a higher-order local model should provide a systematic gain, provided that the derivatives can be computed with sufficient numerical stability.

The error map after translation compensation also deserves separate discussion. The observed regular structure of the error in the presence of periodic textures can be interpreted as a consequence of residual translation or INR approximation errors: in high-frequency regions, even a small error in $\hat{\tau}$ produces a systematic phase mismatch, which manifests as striped patterns. This is important from a practical standpoint because it indicates the limits of purely local criteria: for textured regions, one needs either robust point weighting or a mechanism that explicitly accounts for the signal's frequency properties. In this sense, the results are consistent with recent ideas on processing INR via derivatives as carriers of invariant properties, where local differential operators are viewed as natural tools for continuous signals and as an alternative to explicit decoding onto a grid.

An additional advantage of coordinate-space matching is the localization of discrepancies: the map $|g(x) - f(x+\hat{\tau})|$ makes it possible to identify regions where the error is caused by residual translation or INR approximation error, which is fundamentally harder to obtain from global moment-type aggregates without additional processing.

The comparison with the moment-based baseline exhibits the expected behavior: shift estimation via centroid differences is highly accurate for a controlled integer translation with a stable overlap region, and Hu invariants provide an additional control of translation invariance of shape characteristics. At the same time, moment methods are global aggregates and may lose reliability under partial overlap, changes in brightness/contrast, additive noise, or local deformations, since such factors affect integral characteristics. In the proposed INR-oriented setting, local methods based on partial derivatives can, by contrast, be adapted through region selection and robust point rejection, which is a potential advantage for more challenging scenarios where global moments become unstable. Thus, this comparison should not be interpreted as a competition with the moment approach in the simplest setting, but rather as a demonstration that INR derivatives provide a workable mechanism to estimate a group parameter without transitioning to a discrete representation.

From the standpoint of result credibility, it is important to emphasize that the experimental protocol is controlled: the true translation τ^* is specified explicitly, and the quality of the INR representations is verified via MSE/PSNR, which reduces the risk of interpretational artifacts. However, the accuracy of the estimate $\hat{\tau}$ depends on several factors: the INR approximation quality, the choice of the sampling set X, the proportion of points in textured regions, and the numerical stability of second-order derivatives. In particular, with non-smooth activations, second derivatives may be degenerate or unstable; therefore, for practical second-order use it is advisable to employ smooth

INR architectures. Another limitation is that the proposed approach is currently tailored to pure translations; in the presence of rotation, scaling, or perspective distortions, the group model and the local approximation must be extended.

The practical relevance of the proposed approach lies in enabling alignment and consistency validation in systems where data are already represented as INR: neural fields, neural radiance fields, implicit texture fields, and other scenarios in which preserving continuity and scale-independence is critical. Such a mechanism can be used as a module for pre-normalization, reconstruction validation, or as a component of a broader registration procedure.

Further research is warranted in several directions. First, the formulation should be extended to broader group actions (rotations, affine transformations) and the higher-order local models should be aligned with the corresponding group geometry. Second, robust schemes for selecting the sampling points X (including those that account for spectral properties and local gradient informativeness) should be investigated, along with the impact of noise and incomplete overlap. Third, combining local derivative-based criteria with invariant aggregates appears promising, potentially yielding hybrid methods that can operate in both global and local regimes while remaining compatible with the continuous nature of INR.

Conclusions

In this work, we solve the scientific problem of establishing the translational equivalence of two images represented as implicit neural representations, without decoding them onto a pixel grid. We propose a procedure for estimating the translation parameter in the continuous coordinate space that is consistent with the group formulation $T(2)$ and relies on INR values and their partial derivatives.

Most important scientific results:

1. For the first time, a formalized criterion of translational equivalence for INR images in the coordinate space is obtained. It is based on a local model of the translation group action and on the use of jets (first- and second-order derivatives) as available differential characteristics. Unlike existing approaches, the proposed test of equivalence and the estimation of the group-action parameter are performed without switching to a discrete representation, which *enables* preserving the continuous nature of the model and exploiting analytically computable derivatives.
2. An improved local model for translation estimation is developed by incorporating second-order partial derivatives, which *enables* reducing the linearization error for medium and large translations compared to the first-order method and improving the accuracy of translation-parameter estimation in the presence of high-frequency signal components.
3. Further developed is the representation of translation consistency via compensation in the coordinate space, which provides a visual and quantitative quality control over the overlap region and *enables* localizing regions of systematic mismatch caused by residual translation, INR approximation error, or the signal structure.

Most important practical results and quantitative indicators. The algorithm is experimentally validated on the standard grayscale *Cameraman* image under a controlled translation setting with a padded frame. It is shown that the use of second derivatives leads to a noticeable reduction in translation estimation error compared with the use of only first derivatives, and that this improvement increases as$\| \tau^* \|_2$ grows, as evidenced by the aggregated error-versus-shift dependencies. A moment-based baseline using the centroid and Hu invariants is also included as a control, demonstrating the translation invariance of the corresponding integral characteristics and serving as a reference in the simple pure-translation setting. Across a series of controlled translations, the second-order refinement reduces the translation estimation error on average by approximately a factor of 2-3 compared with the first-order method, while the exact quantitative values are determined by the complete experimental protocol.

Practical significance and recommendations for use. The obtained results are applicable in scenarios where images or fields are already represented as implicit neural representations, also known as neural fields. This includes automated consistency checking of reconstructions, preliminary alignment or normalization prior to further processing, correctness control in field fusion, and data preparation in computer vision tasks. It is recommended to use the first-order method as a fast initialization and the second-order method as a refinement for larger translations or textured regions. For stable use of the second-order method, smooth INR architectures should be selected and the approximation quality of INR should be sufficiently high, as confirmed by MSE and PSNR metrics.

Further development should focus on extending the approach to broader transformation groups such as rotations, affine transformations, and projective transformations with corresponding higher-order local models. Another important direction is the design of robust strategies for selecting sampling points X and weighting schemes that take into account local gradient informativeness and spectral properties of the signal. It is also necessary to study the impact of noise, partial overlap, and photometric differences. Finally, combining derivative-based criteria with invariant aggregates is a promising direction for constructing hybrid alignment procedures that remain compatible with the continuous nature of implicit neural representations.

ADDITIONAL INFORMATION

DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS

In preparing this work, the author used ChatGPT for grammar and spelling checks, paraphrasing, rephrasing, and translation into English. After using these tools, the author reviewed and edited the content and takes full responsibility for the content of this publication.

1. Kuglin C. D. The phase correlation image alignment method / C. D. Kuglin, D. C. Hines // Proceedings of the 1975 IEEE International Conference on Cybernetics and Society. – 1975. – P. 163–165.

2. Guizar-Sicairos M. Efficient subpixel image registration algorithms / M. Guizar-Sicairos, S. T. Thurman, J. R. Fienup // Optics Letters. – 2008. – Vol. 33, № 2. – P. 156–158. – URL: https://doi.org/10.1364/OL.33.000156.

3. Lucas B. D. An iterative image registration technique with an application to stereo vision / B. D. Lucas, T. Kanade // Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI). – 1981. – P. 674–679.. – URL: https://hal.science/hal-03697340.

4. Horn B. K. P. Determining optical flow / B. K. P. Horn, B. G. Schunck // Artificial Intelligence. – 1981. – Vol. 17, № 1–3. – P. 185–203. – URL: https://doi.org/10.1016/0004-3702(81)90024-2.

5. Hu M.-K. Visual pattern recognition by moment invariants / M.-K. Hu // IRE Transactions on Information Theory. – 1962. – Vol. 8, № 2. – P. 179–187. – URL: https://doi.org/10.1109/TIT.1962.1057692.

6. Bedratyuk L., Bedratiuk A. Quality analysis of image scaling methods using moment invariants / L. Bedratyuk, A. Bedratiuk // Measuring and Computing Devices in Technological Processes. – 2020. – No. 1. – P. 51–59.– URL: https://doi.org/10.31891/2219-9365-2020-65-1-9

7. Sarvaiya J. Image registration using NSCT and invariant moment / J. Sarvaiya, S. Patnaik, H. Goklani // International Journal of Image Processing (IJIP). – 2010. – Vol. 4, № 2. – P. 119–130.

8. Sitzmann V., Martel J., Bergman A., Lindell D., Wetzstein G. Implicit neural representations with periodic activation functions / V. Sitzmann et al. // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 7462–7473.– URL: https://doi.org/10.48550/arXiv.2006.09661.

9. Tancik M., Srinivasan P. P., Mildenhall B., et al. Fourier features let networks learn high frequency functions in low dimensional domains / M. Tancik et al. // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – URL: https://doi.org/10.48550/arXiv.2006.10739.

10. Xu D., Wang P., Jiang Y., Fan Z., Wang Z. Signal processing for implicit neural representations / D. Xu et al. // Advances in Neural Information Processing Systems. – 2022, 35, P.13404-13418.– URL: https://doi.org/10.48550/arXiv.2210.08772.

11. Higgins I., Matthey L., Pal A., et al. beta-VAE: Learning basic visual concepts with a constrained variational framework / I. Higgins et al. // International Conference on Learning Representations. – 2017. – URL: https://openreview.net/forum?id=Sy2fzU9gl.

12. Nasiri A. Unsupervised object representation learning using translation and rotation group equivariant VAE / A. Nasiri, T. Bepler // Advances in Neural Information Processing Systems (NeurIPS). – 2022.- Vol.35. – P. 15255-15267. – URL: https://doi.org/10.48550/arXiv.2211.08581.

13. Kwon S. Rotation and Translation Invariant Representation Learning with Implicit Neural Representations / S. Kwon, J. Y. Choi, E. K. Ryu // International Conference on Machine Learning (ICML). – 2023. - P. 18037-18056. – URL: https://doi.org/10.48550/arXiv.2210.08772

Ганна БЕДРАТЮК
Хмельницький національний університет

ОЦІНЮВАННЯ ПАРАМЕТРА ЗСУВУ ДЛЯ НЕЯВНИХ НЕЙРОННИХ РЕПРЕЗЕНТАЦІЙ ЗОБРАЖЕНЬ

Неявні нейронні репрезентації моделюють зображення як неперервні функції координат, що дає змогу виконувати обробку сигналів без дискретизації. Водночас перевірка того, чи відрізняються два зображення лише паралельним зсувом, як правило здійснюється методами, які потребують дискретного подання або глобальних агрегувань, що ускладнює їх пряме застосування до неявних моделей і суперечить їхній неперервній природі. Метою роботи є запропонувати та обґрунтувати алгоритм, який визначає, чи пов'язані дві неявні нейронні репрезентації зображення паралельним зсувом, використовуючи лише значення похідних до другого порядку, а також забезпечити оцінювання параметра зсуву. Запропонований підхід ґрунтується на локальній лінеаризації оператора зсуву та використанні аналітично доступних похідних, отриманих за допомогою автоматичного диференціювання в межах неявної моделі. Критерій узгодженості будується на перевірці стабільності оціненого зсуву на множині точок області визначення. Для підвищення надійності локальні оцінки агрегуються з використанням робастних процедур разом із перевірками узгодженості, що зменшують вплив неоднорідних ділянок сигналу. Розроблено критерій виявлення зсувної еквівалентності двох неявних репрезентацій та процедуру оцінювання параметра зсуву. Показано, що запропонований критерій узгоджується з неперервною природою неявних моделей, не потребує

декодування у піксельну ґратку та є придатним для застосування до моделей, у яких похідні доступні аналітично. Запропонований підхід має наукову новизну у вигляді тесту зсувної еквівалентності для неявних нейронних репрезентацій та практичну значущість як інструмент для швидкої перевірки узгодженості, валідації та попередньої нормалізації даних у задачах комп'ютерного бачення.

Отримані результати застосовні в сценаріях, де зображення або поля вже представлені як неявні нейронні представлення, також відомі як нейронні поля. Це включає автоматизовану перевірку узгодженості реконструкцій, попереднє вирівнювання або нормалізацію перед подальшою обробкою, контроль правильності об'єднання полів та підготовку даних у завданнях комп'ютерного зору. Рекомендується використовувати метод першого порядку як швидку ініціалізацію, а метод другого порядку як уточнення для більших переміщень або текстурованих областей. Для стабільного використання методу другого порядку слід вибирати гладкі архітектури INR, а якість апроксимації INR повинна бути достатньо високою, що підтверджено метриками MSE та PSNR.

Ключові слова: неявні нейронні репрезентації, зсувна еквівалентність, автоматичне диференціювання, оцінювання зсуву, інваріантність до трансляції, аналіз сигналів.