

Quantifying Uncertainty of Image Labelings Using Assignment Flows

Daniel Gonzalez-Alvarado^{1,2}[0000-0002-4636-3697], Alexander
Zeilmann¹[0000-0002-8119-0349], and Christoph Schnörr^{1,2}[0000-0002-8999-2338]

¹ Image and Pattern Analysis Group, Heidelberg University, Germany

² Heidelberg Collaboratory for Image Processing, Heidelberg University, Germany

Abstract. This paper introduces a novel approach to uncertainty quantification of image labelings determined by assignment flows. Local uncertainties caused by ambiguous data and noise are estimated by fitting Dirichlet distributions and pushed forward to the tangent space. The resulting first- and second-order moments are then propagated using a linear ODE parametrization of assignment flows. The corresponding moment evolution equations can be solved in closed form and numerically evaluated using iterative Krylov subspace techniques and low-rank approximation. This results in a faithful representation and quantification of uncertainty in the output space of image labelings, which is important in all applications where confidence in pixelwise decisions matters.

Keywords: image labeling · uncertainty quantification · assignment flows.

1 Introduction

1.1 Overview, Motivation

Quantifying the uncertainty of image segmentations and labelings is an important topic in the field of image analysis and computer vision. For more than three decades, probabilistic graphical models [35] and Bayesian inference was the method of choice. Due to the intractability of exact inference, however, quantifying uncertainty of inference, e.g. by evaluating marginal probabilities of the posterior distribution, requires either computationally expensive MCMC-based sampling methods or variational approximations [9] whose performance is difficult to assess. Another line of research relying on continuous variational approaches to image segmentation, like the relaxed Mumford-Shah functional [3], employs polynomial chaos expansions for uncertainty propagation [22], a framework for approximate stochastic computing that is widely used in scientific computing [36]. The recent work [7] focuses on binary image segmentation from the viewpoint of semi-supervised learning using an approach that combines ideas from phase-field models [3], classical Gaussian processes for classification [23] and spectral graph theory [12].

In summary, before the era of deep networks, methods for quantifying the uncertainty of image segmentations mainly stressed the *Bayesian viewpoint* and

employed established methods for *approximate* inference. Accordingly, there are two aspects that put into question the trustworthiness of corresponding uncertainty estimates: On the one hand, uncertainty quantification using methods where exact inference is intractable defines an intractable problem as well, and quantifying the approximation error is a hard task. On the other hand, Bayesian inference itself has been increasingly criticized recently as being too sensitive to misspecification of models and priors [20,13].

For several years, the state of the art in image segmentation is based on deep networks [19]. It is well-known too, however, that current deep network architectures come along with deficiencies: sensitivity against tiny perturbations (e.g. adversarial attacks) with unpredictable consequences, and with insufficient theoretical understanding of how predictions are generated. It is not surprising, therefore, that uncertainty quantification based on deep networks is an unsolved problem that has been tackled by hundreds of different architectures and heuristics during the last years – see [1] and more than 700 references therein.

1.2 Contribution, Organization

This paper is based on the *assignment flow* approach to image labeling [4,30] that provides a framework for studying the mathematics of deep networks in ‘small controlled’ steps, yet outperforms already in its present form traditional variational and graphical models (cf., e.g. [31,32]). Specifically, we consider the *linearized assignment flow* [38] which enjoys a parametrization through a *linear ODE* (ordinary differential equation) on the tangent space and reasonably approximates the (full) assignment flow [38,37].

We employ a basic method for estimating *locally* initial first- and second-order statistical moments from given input data. Next, we exploit the specific structure of the linearized assignment flow for propagating these initial uncertainties to the *nonlocal* labelings that result from integrating numerically the flow.

Our approach should not be confused with ‘normalizing flows’ [18,21] that estimate a generative model of the data distribution through optimal transport of a reference distribution to a given empirical distribution (data). Rather, we *quantify* the *uncertainty of label assignments* performed by a context-sensitive nonlocal process (linearized assignment flow), caused by uncertainties of noisy and ambiguous initial data.

Section 2 collects basic material required for presenting our contribution in Section 3. Details of our implementation are specified in Section 4. Proof-of-concept experiments validate and illustrate our approach in Section 5. We conclude in Section 6.

2 Preliminaries

2.1 Categorical Distributions, Dirichlet Distribution

We denote in this paper by c the number of classes (labels, categories). *Categorical distributions* are points $p \in \Delta_{c-1} = \{q \in \mathbb{R}_+^c : \langle \mathbb{1}_c, q \rangle = 1\}$ in the probability

simplex. The subset of strictly positive vectors

$$\mathcal{S} = \{p \in \Delta_{c-1} : p_1 > 0, \dots, p_c > 0\} \quad (1)$$

together with the Fisher-Rao metric $g_p(u, v) = \langle u, \text{Diag}(p)^{-1}v \rangle$, $u, v \in T_0$ [2] becomes a Riemannian manifold (\mathcal{S}, g) with trivial tangent bundle $T\mathcal{S} = \mathcal{S} \times T_0$ and tangent space $T_0 = \{v \in \mathbb{R}^c : \langle \mathbb{1}_c, v \rangle = 0\}$. The orthogonal projection onto T_0 is denoted by

$$\Pi_0: \mathbb{R}^c \rightarrow T_0, \quad v \mapsto v - \langle \mathbb{1}_c, v \rangle \mathbb{1}_c, \quad \mathbb{1}_c = \frac{1}{c} \mathbb{1}_c. \quad (2)$$

The *Dirichlet distribution* [17,8]

$$\mathcal{D}_\alpha(p) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_c)} p_1^{\alpha_1-1} \cdots p_c^{\alpha_c-1} \mathbb{1}_\mathcal{S}(p), \quad \alpha \in \mathbb{R}_{>0}^c, \quad \alpha_0 = \langle \mathbb{1}_c, \alpha \rangle \quad (3)$$

with $\mathbb{1}_\mathcal{S}(p) = 1$ if $p \in \mathcal{S}$ and 0 otherwise, belongs to the exponential family of distributions [5,35]. It is strictly unimodal as long as the concentration parameters satisfy $\alpha_1, \dots, \alpha_c \geq 1$. We denote by $\mathbb{E}_\alpha[\cdot]$, $\text{Cov}_\alpha[\cdot]$ the expectation and covariance operator with respect to \mathcal{D}_α and record the relations for a random vector $p \sim \mathcal{D}_\alpha$

$$\mathbb{E}_\alpha[\log p] = \psi(\alpha) - \psi(\alpha_0) \mathbb{1}_c, \quad (4a)$$

$$\text{Cov}_\alpha[\log p, \log p] = \text{Diag}(\psi'(\alpha)) - \psi'(\alpha_0) \mathbb{1}_c \mathbb{1}_c^\top, \quad (4b)$$

where $\psi(\alpha) := (\psi(\alpha_1), \dots, \psi(\alpha_c))^\top$ and ψ, ψ' are the digamma and trigamma functions, respectively [16, pp. 8–9].

We will use Dirichlet distributions in Section 3.1 for estimating local uncertainties of input data, to be propagated to uncertainties of non-local labelings by linearized assignment flows in Sections 3.2 and 3.3.

2.2 Linearized Assignment Flows

Linearized assignment flows (LAFs) were introduced by [38] as approximations of (full) assignment flows [4,30] for metric data labeling on graphs $G = (I, E)$. The linearization concerns the parametrization of assignment flows on the tangent space T_0 through a *linear* ODE with respect to a linearization point $W_0 \in \mathcal{W} := \mathcal{S} \times \cdots \times \mathcal{S}$ ($n = |I|$ factors). Each point $W \in \mathcal{W} \subset \mathbb{R}^{n \times c}$ of the assignment manifold \mathcal{W} comprises as row vector a categorical distribution $W_i \in \mathcal{S}$, $i \in I$, with \mathcal{S} defined by (1). Adopting the barycenter $W_0 = \mathbb{1}_\mathcal{W}$ of the assignment manifold defined by $(\mathbb{1}_\mathcal{W})_i = \mathbb{1}_\mathcal{S}$ (cf. (2)) as simplest choice of the point of linearization, the LAF takes the form

$$\dot{V} = V_D + \Omega V, \quad V_D = \Pi_0 L_D, \quad V(0) = 0, \quad (5)$$

where $V(t) \in \mathcal{T}_0 = T_0 \times \cdots \times T_0 \subset \mathbb{R}^{n \times c}$ ($|I|$ factors), i.e. each row vector $V_i(t)$ evolves on T_0 , $\Omega \in \mathbb{R}_+^{n \times n}$ is a parameter matrix that defines the regularization

properties of the LAF, and

$$L_D \in \mathcal{W}, \quad L_{D;i} = \exp_{\mathbb{1}_S}(-D_i/\rho) := \frac{e^{-D_i/\rho}}{\langle \mathbb{1}_c, e^{-D_i/\rho} \rangle}, \quad \rho > 0, \quad i \in I \quad (6)$$

encodes the input data as point on the assignment manifold (here, the exponential function applies componentwise). The input data are given as distance vector field $D \in \mathbb{R}^{n \times c}$, where each vector $D_i \in \mathbb{R}^c$ encodes the distances D_{ij} , $j \in [c]$ of the datum (feature) observed at vertex $i \in I$ to c pre-specified labels (class prototypes). Both integration (inference) of (5) and supervised learning of the parameter matrix Ω by minimizing a LAF-constrained loss function can be computed efficiently [38,37]. Once $V(T)$ has been computed for a sufficiently large time $t = T$, almost hard label assignments to the data are given by the assignment vectors

$$W(T), \quad W_i(T) = \exp_{\mathbb{1}_S}(V_i(T)) = \frac{e^{V_i(T)}}{\langle \mathbb{1}_c, e^{V_i(T)} \rangle}, \quad i \in I. \quad (7)$$

Our contribution in this paper to be developed subsequently is to augment labelings (7) with uncertainty estimates, by propagating initial local uncertainties that are estimated from the data using the LAF.

3 Modeling and Propagating Uncertainties

3.1 Data-Based Local Uncertainty Estimation

Local initial labeling information is given by the input data (6) and the weight patches

$$\Omega_i = \{\omega_{ik} : k \in \mathcal{N}_i\}, \quad \mathcal{N}_i = \{i\} \cup \{k \in I : i \sim k\}, \quad i \in I, \quad (8)$$

that form sparse row vectors of the weight matrix Ω of the LAF equation (5). We denote the weighted geometric averaging of the categorical distributions (6) by

$$S_i := \Omega_i * L_D := \frac{\prod_{k \in \mathcal{N}_i} (L_{D;k})^{w_{ik}}}{\langle \mathbb{1}_c, \prod_{k \in \mathcal{N}_i} (L_{D;k})^{w_{ik}} \rangle}, \quad i \in I, \quad (9)$$

where exponentiation is done componentwise. This formula can be derived by performing a single step towards the Riemannian center of mass of the vectors (6), but with the e-connection from information geometry in place of the Riemannian (Levi-Civita) connection [30, Section 2.2.2].

Local data variations can thus be represented at every pixel $i \in I$ by the set

$$\mathcal{F}_i := \{S_j : j \in \mathcal{N}_i\}, \quad \text{where } S_j = \Omega_j * L_D. \quad (10)$$

We use Dirichlet distributions as a natural statistical model to represent the initial uncertainties of these data, i.e. $\mathcal{F}_i \sim \mathcal{D}_{\alpha^i}$ at every pixel $i \in I$. We estimate

the parameters α^i by maximizing the corresponding log-likelihood. To this end, we exploit the structure of Dirichlet distributions as member of the exponential family and rewrite the densities (3) in the form

$$\mathcal{D}_{\alpha^i} : S \mapsto e^{\langle T(S), F(\alpha^i) \rangle - \zeta(\alpha^i)}, \quad i \in I, \quad (11)$$

where

$$T(S) = \log(S), \quad F(\alpha^i) = \alpha^i - \mathbb{1}_c, \quad \zeta(\alpha^i) = \sum_{k=1}^c \log \Gamma(\alpha_k^i) - \log \Gamma(\alpha_0^i). \quad (12)$$

Assuming that $S_1, \dots, S_{|\mathcal{N}_i|}$ are i.i.d. samples at every pixel $i \in I$, for simplicity, it follows that $S_{\mathcal{F}_i} := \sum_{j \in \mathcal{N}_i} T(S_j)$ is a sufficient statistic for the Dirichlet class \mathcal{D}_{α^i} [10, Theorem 1.11]. Hence the parameters α^i which best fit the data \mathcal{F}_i can be obtained by maximizing the log-likelihood functions

$$\mathcal{L}_i : \mathbb{R}_{>0}^c \rightarrow \mathbb{R}, \quad \alpha \mapsto \mathcal{L}_i(\alpha) = \langle S_{\mathcal{F}_i}, F(\alpha) \rangle - |\mathcal{N}_i| \zeta(\alpha), \quad i \in I. \quad (13)$$

This is a concave optimization problem [25], [10, Lemma 5.3] with derivatives of the objective function given by

$$\nabla \mathcal{L}_i(\alpha) = |\mathcal{N}_i| \left(\psi(\alpha_0) \mathbb{1}_c - \psi(\alpha) + \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \log(S_j) \right), \quad (14a)$$

$$\mathcal{H}(\mathcal{L}_i)(\alpha) = |\mathcal{N}_i| \left(-\text{Diag}(\psi'(\alpha)) + \psi'(\alpha_0) \mathbb{1}_c \mathbb{1}_c^\top \right). \quad (14b)$$

We solve this problem numerically as specified in Section 4 and point here merely out that Hessian-based algorithms are convenient to apply due to the structure of (14a), since $\mathcal{H}(\mathcal{L}_i)^{-1}(\alpha)$ can be computed in linear time [15].

3.2 Local Uncertainty Representation

We transfer the local data uncertainties obtained in the preceding section in terms of the Dirichlet distributions \mathcal{D}_{α^i} , $i \in I$ to the tangent space T_0 , in order to compute first- and second-order moments. These moments will be propagated to labelings using the LAF in Section 3.3.

The following proposition uses the inverse of the diffeomorphism $\exp_{\mathbb{1}_S}$ defined by (6),

$$\phi : \mathcal{S} \rightarrow T_0, \quad p \mapsto \phi(p) := \exp_{\mathbb{1}_S}^{-1}(p) = \Pi_0 \log p, \quad (15)$$

where the log applies componentwise. See [29, Lemma 3.1] for more details.

Proposition 1. *Let \mathcal{D}_α be a given Dirichlet distribution, and let $\nu = \phi_\# \mathcal{D}_\alpha$ denote the pushforward measure of \mathcal{D}_α by ϕ and $\mathbb{E}_\nu, \text{Cov}_\nu$ the corresponding expectation and covariance operators. Then, for a random vector $v \sim \nu$,*

$$\mathbb{E}_\nu[v] = \Pi_0 \psi(\alpha), \quad (16a)$$

$$\text{Cov}_\nu[v, v] = \Pi_0 \text{Diag}(\psi'(\alpha)) \Pi_0. \quad (16b)$$

Proof. By definition of the pushforward operation, for any integrable function $f: T_0 \rightarrow \mathbb{R}$ and $v = \phi(p)$, one has $\mathbb{E}_\nu[f(v)] = \mathbb{E}_\alpha[f \circ \phi(p)]$ and consequently

$$\mathbb{E}_\nu[v_j] = \mathbb{E}_\alpha[\phi(p)_j] \stackrel{(15),(2)}{=} \mathbb{E}_\alpha[\log p_j - \langle \mathbb{1}_S, \log p \rangle] \stackrel{(4a)}{=} \psi(\alpha_j) - \langle \mathbb{1}_S, \psi(\alpha) \rangle, \quad (17)$$

which implies (16a). Regarding (16b), we compute

$$\text{Cov}_\nu[v, v] = \mathbb{E}_\nu[(v - \mathbb{E}_\nu[v])(v - \mathbb{E}_\nu[v])^\top] = \Pi_0 \text{Cov}_\alpha[\log p, \log p] \Pi_0 \quad (18)$$

which implies (16b) by (4b) and by the equation $\Pi_0 \mathbb{1}_c = 0$. \square

Applying this proposition with $V_i(0)$ and α^i in place of v and α , respectively, we get the first- and second-order Dirichlet moments lifted to T_0 ,

$$m_{0,i} := \Pi_0 \psi(\alpha^i), \quad (19a)$$

$$\Sigma_{0,i} := \Pi_0 \text{Diag}(\psi'(\alpha^i)) \Pi_0. \quad (19b)$$

3.3 Uncertainty Propagation

In this section, we model the initial data of the LAF (5) as Gaussian Markov random field (GMRF) and study the Gaussian random process induced by the LAF, as a model for propagating the initial data uncertainties to the resulting labeling.

Given the lifted Dirichlet moments (19), we regard each initial vector $V_i(0)$, $i \in I$ of the LAF (5) as normally distributed vector

$$V_i(0) \sim \mathcal{N}(m_{0,i}; \Sigma_{0,i}). \quad (20)$$

The formulae (19) show that these normal distributions are supported on the tangent space T_0 .

Next, we stack the row vectors $V_i(0)$, $i \in I$ of $V(0)$ and denote the resulting vector by

$$v(0) := \text{vec}_r(V(0)) := (V_1(0)^\top, V_2(0)^\top, \dots, V_n(0)^\top)^\top, \quad n = |I|. \quad (21)$$

Thus, $v(0)$ is governed by the GMRF

$$v(0) \sim \mathcal{N}(m_0, \Sigma_0), \quad m_0 = \text{vec}_r(m_{0,i})_{i \in I}, \quad \Sigma_0 = \text{BlockDiag}(\Sigma_{0,i})_{i \in I}. \quad (22)$$

In fact, since the covariance matrices $\Sigma_{0,i}$ are singular with rank $c - 1$, this is an *intrinsic* GMRF of first order – cf. [26, Chapter 3]. Conforming to (21), we transfer the LAF equation to its vectorized version

$$\dot{v} = v_D + \Omega_c v, \quad v(0) \sim \mathcal{N}(m_0, \Sigma_0), \quad (23)$$

where $\dot{v} = \text{vec}_r(\dot{V})$, $v = \text{vec}_r(V)$, $v_D = \text{vec}_r(V_D)$ and $\Omega_c = \Omega \otimes I_c$ using the stacking convention $\text{vec}_r(ABC) = (A \otimes C^\top) \text{vec}_r(B)$ using the Kronecker product \otimes [34] for arbitrary matrices A, B, C with compatible dimensions.

The evolution of the initial probability distribution that governs the evolving state $v(t)$ of (23) is generally described by the Fokker-Planck equation [24,28]. For the specific simple case considered here, i.e. a linear deterministic ODE with random initial conditions, this is a Gaussian random process with moments determined by the differential equations

$$\dot{m}(t) = \mathbb{E}[\dot{v}(t)], \quad m(0) = m_0 \quad (24a)$$

$$\dot{\Sigma}(t) = \mathbb{E}[\dot{v}(t)(v(t) - m(t))^\top] + \mathbb{E}[(v(t) - m(t))\dot{v}(t)^\top], \quad \Sigma(0) = \Sigma_0, \quad (24b)$$

with initial moments given by (23). Equations (24) propagate the uncertainty on the tangent space until the time of point $t = T$ and to the labeling in terms of the distribution governing the state $v(T) = \text{vec}_r(V(T)) \sim \mathcal{N}(m(T), \Sigma(T))$ and equation (7). Regarding the quantitative evaluation of these uncertainties, we focus on the *marginal* distributions of the subvectors $V_i(t)$, $i \in I$.

Proposition 2. *Let $v(t)$ be the random vector solving (23). Then the first- and second-order moments of the marginal distributions of $V_i(t) = (\text{vec}_r^{-1}(v(t)))_i$ are given by*

$$(m_i(t))_{i \in I} = m(t) = \text{expm}(t\Omega_c)m_0 + t\varphi(t\Omega_c)v_D \quad (25a)$$

$$\Sigma_i(t) = \sum_{j \in I} \left((\text{expm}(t\Omega_c))_{ij} \right)^2 \Sigma_{0,j}, \quad i \in I, \quad (25b)$$

where expm denotes the matrix exponential and φ the matrix-valued function given by the entire function $\varphi : z \mapsto \frac{e^z - 1}{z}$.

Proof. The solution (25a) results from the linear equation

$$\dot{m}(t) = \mathbb{E}[\dot{v}(t)] = \mathbb{E}[\Omega_c v(t) + v_D] = \Omega_c m(t) + v_D \quad (26)$$

and applying Duhamel's formula [33, Eq. (3.48)], to obtain

$$m(t) = \text{expm}(t\Omega_c)m_0 + \text{expm}(t\Omega_c) \int_0^t \text{expm}(-s\Omega_c)v_D ds \quad (27a)$$

$$= \text{expm}(t\Omega_c)m_0 + t \text{expm}(t\Omega_c) \sum_{k=0}^{\infty} \left[\frac{(-t\Omega_c)^k}{(k+1)!} \right] v_D \quad (27b)$$

$$= \text{expm}(t\Omega_c)m_0 + t\varphi(t\Omega_c)v_D. \quad (27c)$$

In order to derive (25b), we first compute separately the two terms on the r.h.s. of (24b) which yields

$$\dot{\Sigma}(t) = \Omega_c \Sigma(t) + \Sigma(t) \Omega_c^\top. \quad (28)$$

This differential Lyapunov equation has the solution [6, Theorem 1]

$$\Sigma(t) = \text{expm}(t\Omega_c) \Sigma_0 (\text{expm}(t\Omega_c))^\top. \quad (29)$$

In order to compute the marginal covariance matrices, we use properties of the Kronecker product [34] and compute

$$\expm((t\Omega_c)) = \sum_{k=0}^{\infty} \frac{(t\Omega)^k \otimes (I_c)^k}{k!} = \sum_{k=0}^{\infty} \left(\frac{(t\Omega)^k}{k!} \right) \otimes I_c = \expm(t\Omega) \otimes I_c, \quad (30)$$

which together with $\expm(t(\Omega \otimes I_c))^\top = \expm(t(\Omega^\top \otimes I_c))$ transforms (29) into

$$\Sigma(t) = (\expm(t\Omega) \otimes I_c) \Sigma_0 (\expm(t\Omega^\top) \otimes I_c). \quad (31)$$

Since the matrix $\Sigma(0) = \Sigma_0$ is block diagonal, we have

$$(\expm(t\Omega) \otimes I_c) \Sigma_0 = \begin{pmatrix} \expm(t\Omega)_{1,1} \Sigma_{0,1} & \cdots & \expm(t\Omega)_{1,n} \Sigma_{0,n} \\ \vdots & \ddots & \vdots \\ \expm(t\Omega)_{n,1} \Sigma_{0,1} & \cdots & \expm(t\Omega)_{n,n} \Sigma_{0,n} \end{pmatrix}, \quad n = |I| \quad (32)$$

and thus can extract from (31) the covariance matrices

$$\Sigma_i(t) = \sum_{j \in I} (\expm(t\Omega))_{ij} \Sigma_{0,j} (\expm(t\Omega^\top))_{ji} = \sum_{j \in I} \left((\expm(t\Omega))_{ij} \right)^2 \Sigma_{0,j} \quad (33)$$

□

of the marginal distributions. The direct numerical evaluation of the equations (25) is infeasible for typical problem sizes, but can be conveniently done using a low-rank approximation; see Section 4.2.

4 Algorithms

4.1 Estimating Local Uncertainties and Moments

We use gradient ascent to maximize numerically the log-likelihoods \mathcal{L}_i in (13) for every pixel $i \in I$. The corresponding global maxima correspond to parameter values α^i which best represent local data variations S_j in (10) in terms of Dirichlet distributions \mathcal{D}_{α^i} given by (3) and (11), respectively. While numerical gradient ascent is safe, Hessian-based optimization method (cf. (14a)) may be more time-efficient but require numerical damping to ensure convergence [25,15].

4.2 Moment Evolution

The explicit computation of the moments $m_i(t)$ and $\Sigma_i(t)$ of the marginal uncertainty distributions in (25) require the evaluation of the matrix exponential followed by a matrix-vector multiplication. Such operations are typically intractable numerically due to the large size of the involved matrices. Recall that the size of $\Omega \in \mathbb{R}^{|I| \times |I|}$ is quadratic in the number of pixels. We overcome this numerical problem by using low-rank Krylov subspace approximation [27,14] that has shown to perform very well in connection with the linearized assignment flow [38,37].

Evolution of means. We first describe the numerical evaluation of the mean $m_0 := m(0)$ in (25a) at any pixel $i \in I$. Using the Arnoldi iteration, we compute an orthonormal basis for each of the Krylov subspaces

$$\mathcal{K}_d(\Omega, m_0) := \text{span}\{m_0, \Omega m_0, \dots, \Omega^{d-1} m_0\}, \quad (34)$$

$$\mathcal{K}_d(\Omega, v_D) := \text{span}\{v_D, \Omega v_D, \dots, \Omega^{d-1} v_D\}, \quad (35)$$

of dimension $d \ll |I|$. This iterative method yields a basis $V_d = (v_1, \dots, v_d)$ together with the upper Hessenberg matrix

$$H_d = V_d^\top \Omega V_d, \quad (36)$$

without having to compute explicitly the right-hand side of (36). We use this Hessenberg matrix, which represents the projection of Ω onto the Krylov subspace, to approximate the action of the matrix exponential [27]

$$\text{expm}(t\Omega)m_0 \approx \|m_0\| V_d \text{expm}(tH_d)e_1, \quad (37)$$

where e_1 denotes the first unit vector. In agreement with [38,37], our experience is that the computational costs are reduced remarkably while still very accurate approximations are obtained even when working with low-dimensional Krylov subspaces ($d \approx 10$).

In order to approximate the second term on the right-hand side of (25a), we consider the matrices H_d, V_d corresponding to the Krylov space (35) and employ the approximation

$$t\varphi(t\Omega)v_D \approx t\|v_D\| V_d \varphi(tH_d)e_1, \quad (38)$$

where the action of the matrix-valued φ -function on the vector e_1 can be derived using an evaluation of the matrix exponential of the form [27, Proposition 2.1]

$$\text{expm}\left(t \begin{pmatrix} H_d & e_1 \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} \text{expm}(tH_d) & t\varphi(tH_d)e_1 \\ 0 & 1 \end{pmatrix}. \quad (39)$$

Evolution of covariance matrices. For the evolution of each covariance matrix (25b), we compute the vectors

$$\sigma_i = \text{expm}(t\Omega^\top)e_i \in \mathbb{R}^{|I|}, \quad i \in I \quad (40)$$

using the Krylov approximation (37) as described above and evaluate the covariance matrices of the marginal distributions by

$$\Sigma_i(t) = \sum_{j \in I} \sigma_{i,j}^2 \Sigma_{0,j}, \quad i \in I. \quad (41)$$

Here, the matrices $\Sigma_{0,j}$ correspond to the initial local uncertainties estimated at pixel $j \in I$, as described in connection with (19b).

5 Experiments and Findings

We illustrate our approach (Section 3) in a number of proof-of-concept experiments. We refer to Section 4 for implementation details and to the figure captions for a description of the experiments and our findings.

Unless otherwise specified, we used the following options in the experiments:

- The linear assignment flow was integrated up to time $t = T := 15$. This suffices to obtain almost integral assignments by (7).
- All Krylov dimensions (cf. (34), (35)) were set to $d = 10$.
- We used $|\mathcal{N}_i| = 3 \times 3$ neighborhoods for specifying the regularizing parameters (weights) by (8).
- The weight matrix Ω with sparse row vectors due to (8) were either taken as uniform weights or computed using a nonlocal means algorithm [11].

Computation of Uncertainties. The *uncertainty* of assigning a label to the data observed at pixel $i \in I$ is the probability that the random vector $V_i \sim \mathcal{N}(m_i(T), \Sigma_i(T))$ yields a *different* label assignment by (7), where $m_i(T), \Sigma_i(T)$ are given by (25) at time $t = T$. In practice, we compute these uncertainties as normalized frequencies of label confusion using a sufficiently large number of samples $V_i \sim \mathcal{N}(m_i(T), \Sigma_i(T))$. These computations can be carried out in parallel for each pixel $i \in I$.

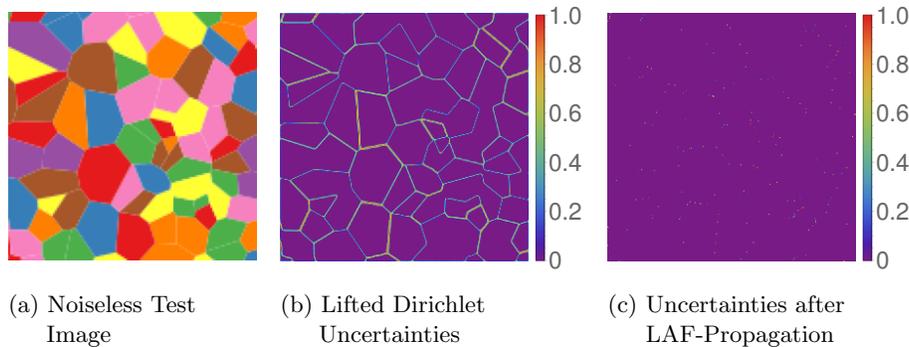


Fig. 1: **Uncertainties for labeling a noiseless image.** (a) Noiseless test image, where each of the eight colors (red, blue, green, purple, orange, yellow, brown, pink) represents a label. (b) Initial uncertainties at $t = 0$ based on the lifted Dirichlet moments (19). Label assignments in the interior of the cells have uncertainty zero, in agreement with the fact that local rounding already produces the correct labeling. The initial decisions at the boundaries between cells are uncertain, however. (c) Uncertainties of final label assignments through the LAF. Regularization performed by the LAF largely removed the initial uncertainties shown by (b), up to pixels in (a) with color values that do *not* correspond to *any* label due to rasterization effects (visible after zooming in on the screen).

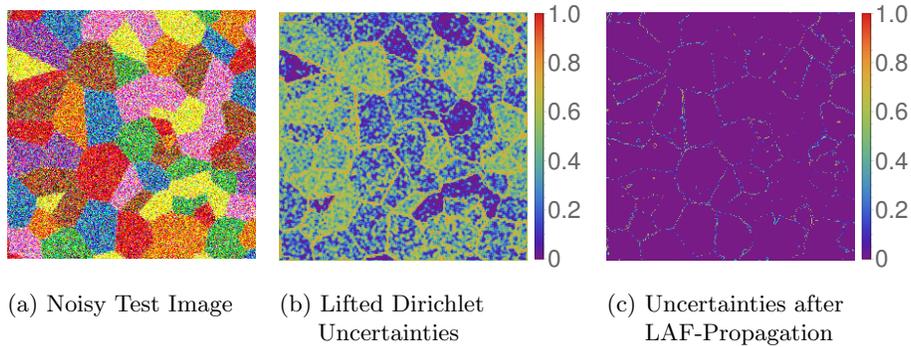


Fig. 2: **Uncertainties for labeling noisy data.** (a) Noisy version of the image from Fig. 1a, (b) Initial uncertainties at $t = 0$ based on the lifted Dirichlet moments (19). Smaller uncertainties in yellow regions reflects the fact that the yellow label has the largest distance to all other labels. (c) Uncertainties of final label assignments through the LAF. The LAF effectively removes noise and feels confident, except for pixels at signal transitions.

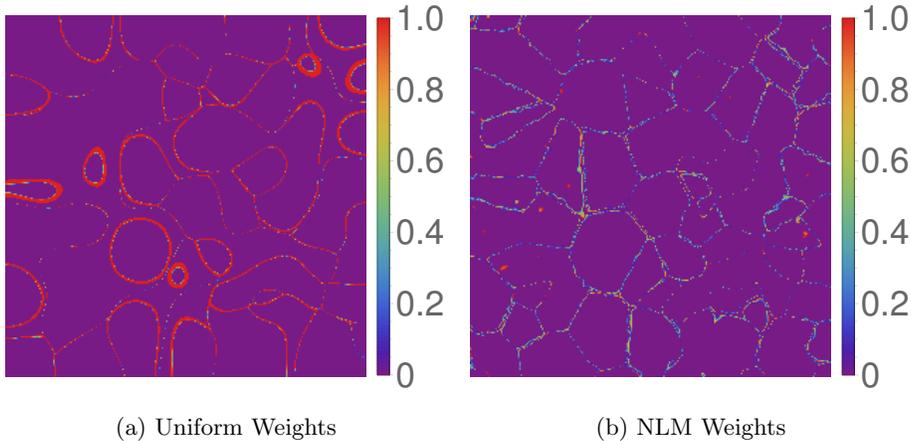


Fig. 3: **Uncertainties for labeling with different weight types.** Both panels display uncertainties of final label assignments through the LAF, obtained using weights computed in two ways, for the noisy data depicted by Figure 2a. (a) Uniform weights size using large patches of size 11×11 pixels, i.e., each weight has the value $\frac{1}{121}$. Such ‘uninformed’ weights (regularization parameters) increase the uncertainty of both label assignments and localization, in particular at small-scale image structures relative to the patch size that determines the strength of regularization of the LAF. (b) Using nonlocal means weights in patches of size 11×11 avoids the effects displayed by (a), despite using the same patch size.

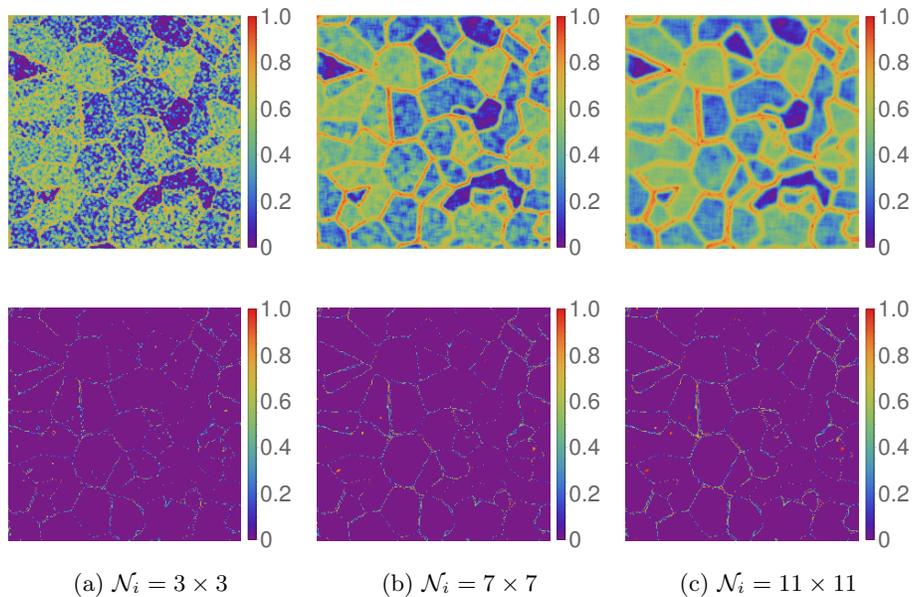


Fig. 4: Uncertainties of labelings using different neighborhood sizes. This figure illustrates the influence of the neighborhood size $|\mathcal{N}_i|$, $i \in I$ on the uncertainties of *initial* label assignments (**top row**) and on the uncertainties of *final* label assignments through the LAF (**bottom row**), for the noisy data depicted by Figure 2a: **(a)** $|\mathcal{N}_i| = 3 \times 3$ (left column), **(b)** $|\mathcal{N}_i| = 7 \times 7$ (center column), and **(c)** $|\mathcal{N}_i| = 11 \times 11$ (right column). Larger neighborhoods increase the initial uncertainties, whereas final uncertainties are fairly insensitive against variation of the neighborhood size due to regularization performed by the LAF.

6 Conclusion and Further Work

We introduced a novel approach to uncertainty quantification of image labelings. The approach is based on simplifying assumptions, like the i.i.d. assumption underlying maximum-likelihood parameter estimation in Section 4.1, regarding the estimation and representation of initial uncertainties of *local* label assignments using the given data. These initial uncertainties are *rigorously* propagated using the mathematical representation of assignment flows. Numerical results validate and illustrate the approach.

Our further work will mainly focus (i) on studying how *learning* regularization parameters of the LAF [37] affects uncertainty quantification, and (ii) on an extension to the *full* assignment flow approach by approximating the flow through a composition of the linearized assignment flow (LAF) determined at few points of time.

Acknowledgement. This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC 2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster).

References

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., R., A.U., Makarenkov, V., Nahavandi, S.: A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. *Information Fusion* (2021, in press)
2. Amari, S.I., Nagaoka, H.: *Methods of Information Geometry*. Amer. Math. Soc. and Oxford Univ. Press (2000)
3. Ambrosio, L., Tortorelli, V.M.: Approximation of Functional Depending on Jumps by Elliptic Functional via Γ -Convergence. *Comm. Pure Appl. Math.* **43**(8), 999–1036 (1990)
4. Åström, F., Petra, S., Schmitzer, B., Schnörr, C.: Image Labeling by Assignment. *Journal of Mathematical Imaging and Vision* **58**(2), 211–238 (2017)
5. Barndorff-Nielsen, O.E.: *Information and Exponential Families in Statistical Theory*. Wiley, Chichester (1978)
6. Behr, M., Benner, P., Heiland, J.: Solution Formulas for Differential Sylvester and Lyapunov Equations. *Calcolo* **56**(4), 1–33 (2019)
7. Bertozzi, A., Luo, X., Stuart, A., Zygalakis, K.: Uncertainty Quantification in Graph-Based Classification of High Dimensional Data. *SIAM/ASA J. Uncertainty Quantification* **6**(2), 568–595 (2018)
8. Bishop, C.: *Pattern Recognition and Machine Learning*. Springer (2006)
9. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* **112**(518), 859–877 (2017)
10. Brown, L.D.: *Fundamentals of Statistical Exponential Families*. Institute of Mathematical Statistics, Hayward, CA (1986)
11. Buades, A., Coll, B., Morel, J.M.: Image Denoising Methods. A New Nonlocal Principle. *SIAM Review* **52**(1), 113–147 (2010)
12. Chung, F.: *Spectral Graph Theory*. Amer. Math. Soc. (1997)
13. Grünwald, P., van Ommen, T.: Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It. *Bayesian Analysis* **12**(4), 1069–1103 (2017)
14. Hochbruck, M., Lubich, C.: On Krylov Subspace Approximations to the Matrix Exponential Operator. *SIAM Journal on Numerical Analysis* **34**(5), 1911–1925 (1997)
15. Huang, J.: *Maximum Likelihood Estimation of Dirichlet Distribution Parameters*. CMU Technique Report (2005)
16. Johnson, N.L., Kemp, A.W., Kotz, S.: *Univariate Discrete Distributions*. Wiley-Interscience, 3rd edn. (2005)
17. Kotz, S., Balakrishnan, N., Johnson, N.L.: *Continuous Multivariate Distributions: Models and Applications*, vol. 1. John Wiley & Sons, Inc., 2nd edn. (2000)
18. Marzouk, Y., Moselhy, T., Parno, M., Spantini, A.: An Introduction to Sampling via Measure Transport. In: *Handbook of Uncertainty Quantification*, pp. 1–41. Springer (2017)

19. Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D.: Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, <https://ieeexplore.ieee.org/document/9356353> (2021)
20. Owahdi, H., Scovel, C.: Brittleness of Bayesian Inference under Finite Information in a Continuous World. *Electr. J. Statistics* **9**, 1–79 (2015)
21. Papamakarios, G., Nalisnick, E., Rezende, D.J. Mohamed, S., Lakshminarayanan, B.: Normalizing Flows for Probabilistic Modeling and Inference. *J. Machine Learning Research* **22**(57), 1–64 (2021)
22. Pätz, T., Kirby, R.M., Preusser, T.: Ambrosio-Tortorelli Segmentation of Stochastic Images: Model Extensions, Theoretical Investigations and Numerical Methods. *Int. J. Computer Vision* **103**, 190–212 (2013)
23. Rasmussen, C.E., Williams, C.: *Gaussian Processes for Machine Learning*. MIT Press (2006)
24. Risken, H.: *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer, 2nd edn. (1989)
25. Ronning, G.: Maximum Likelihood Estimation of Dirichlet Distributions. *Journal of Statistical Computation and Simulation* **32**(4), 215–221 (1989)
26. Rue, H., Held, L.: *Gaussian Markov Random Fields: Theory and Applications*. CRC press (2005)
27. Saad, Y.: Analysis of some Krylov Subspace Approximations to the Matrix Exponential Operator. *SIAM Journal on Numerical Analysis* **29**(1), 209–228 (1992)
28. Särkkä, S., Solin, A.: *Applied Stochastic Differential Equations*, vol. 10. Cambridge University Press (2019)
29. Savarino, F., Schnörr, C.: Continuous-Domain Assignment Flows. *Europ. J. Appl. Math.* **32**(3), 570–597 (2021)
30. Schnörr, C.: Assignment Flows. In: Grohs, P., Holler, M., Weinmann, A. (eds.) *Variational Methods for Nonlinear Geometric Data and Applications*, pp. 235–260. Springer (2020)
31. Sitenko, D., Boll, B., Schnörr, C.: Assignment Flow for Order-Constrained OCT Segmentation. In: *GCPR* (2020)
32. Sitenko, D., Boll, B., Schnörr, C.: Assignment Flow For Order-Constrained OCT Segmentation. *Int. J. Computer Vision* (in press, 2021), <https://link.springer.com/content/pdf/10.1007/s11263-021-01520-5.pdf>
33. Teschl, G.: *Ordinary Differential Equations and Dynamical Systems*, Grad. Studies Math., vol. 140. Amer. Math. Soc. (2012)
34. Van Loan, C.F.: The Ubiquitous Kronecker Product. *J. Comput. Appl. Math.* **123**, 85–100 (2000)
35. Wainwright, M., Jordan, M.: Graphical Models, Exponential Families, and Variational Inference. *Found. Trends Mach. Learn.* **1**(1-2), 1–305 (2008)
36. Xiu, D., Karniadakis, G.E.: The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM J. Sci. Comput.* **24**(2), 619–644 (2002)
37. Zeilmann, A., Petra, S., Schnörr, C.: Learning Linear Assignment Flows for Image Labeling via Exponential Integration. In: Elmoataz, A., Fadili, J., Quéau, Y., Rabin, J., Simon, L. (eds.) *Scale Space and Variational Methods in Computer Vision (SSVM)*. LNCS, vol. 12679, pp. 385–397 (2021)
38. Zeilmann, A., Savarino, F., Petra, S., Schnörr, C.: Geometric Numerical Integration of the Assignment Flow. *Inverse Problems* **36**(3), 034004 (33pp) (2020)