COMPUTATIONAL MATHEMATICS
ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА

Original Theoretical Research

UDC 004.93
https://doi.org/10.23947/2587-8999-2026-10-1-7-20

Semantic Segmentation with Uncertainty Estimation Based on the Dirichlet Model and Anisotropic Regularization

Evgeny Yu. Shchetinin ✉, **Andrey A. Shevchuk**
Sevastopol State University, Sevastopol, Russia
✉ riviera-molto@mail.ru

Abstract
Introduction. In computational mathematics, variational methods for minimizing discrete energies are widely used for the regularization of ill-posed problems. However, standard discrete schemes often suffer from scale inconsistency: upon mesh refinement ($h \to 0$), weights depending on unnormalized jumps of the function degenerate, leading to trivialization of the anisotropic properties of the limiting operator. In this paper, a computational method is proposed that solves this problem by parameterizing the Dirichlet distribution and employing rigorously justified anisotropic spatial regularization.
Materials and Methods. The mathematical model is formulated as an optimization problem for a composite functional in the space of grid functions. The functional includes an expected logarithmic loss function, Kullback-Leibler regularization, and spatial regularizers of the weighted Dirichlet energy type. To ensure the structural consistency of the discrete operator, *edge-aware* weight functions are constructed strictly through normalized finite differences. The asymptotic behavior of the discrete energies is investigated using the apparatus of Γ-convergence.
Results. The main theoretical result of the work is a mathematical proof of the Γ-convergence of a family of discrete anisotropic functionals to a non-trivial continuous limit in the $L^2(\Omega)$ topology. The equicoercivity of the discrete energies is proven, guaranteeing the convergence of a sequence of almost minimizers to the solution of the continuous problem.
Discussion. The use of normalized finite differences in constructing the weights restores dimensional homogeneity and ensures strict scale invariance of the discretization of non-local operators.
Conclusion. The proposed method successfully links continuous variational formulations with discrete predictive models, providing a theoretically justified and computationally efficient tool (additional inference costs amount to 17–18%) with controlled error.

Keywords: semantic segmentation, Dirichlet distribution, uncertainty estimation, probability calibration, anisotropic regularization, Dirichlet energy, Γ-convergence, equicoercivity, medical images

Funding. The work was supported by Sevastopol State University (Project No. 42–01–09/319/2025–1).

For Citation. Shchetinin E.Yu., Shevchuk A.A. Semantic Segmentation with Uncertainty Estimation Based on the Dirichlet Model and Anisotropic Regularization. *Computational Mathematics and Information Technologies*.2026;10(1):7–20. https://doi.org/10.23947/2587-8999-2026-10-1-7-20

Оригинальное теоретическое исследование

Семантическая сегментация с оценкой неопределённости на основе модели Дирихле и анизотропной регуляризации

Е.Ю. Щетинин ✉, **А.А. Шевчук**
Севастопольский государственный университет, г. Севастополь, Российская Федерация
✉ riviera-molto@mail.ru

Аннотация
Введение. В задачах вычислительной математики вариационные методы минимизации дискретных энергий широко применяются для регуляризации некорректных задач. Однако стандартные дискретные схемы зачастую об-

ладают масштабной несогласованностью: при измельчении сетки ($h \to 0$) веса, зависящие от ненормированных скачков функции, вырождаются, что приводит к тривиализации анизотропных свойств предельного оператора. В данной работе предложен вычислительный метод, решающий эту проблему за счет параметризации распределения Дирихле и строго обоснованной анизотропной пространственной регуляризации.
Материалы и методы. Математическая модель формулируется как задача оптимизации составного функционала в пространстве сеточных функций. Функционал включает ожидаемую логарифмическую функцию потерь, регуляризацию Кульбака-Лейблера и пространственные регуляризаторы типа взвешенной энергии Дирихле. Для обеспечения структурной состоятельности дискретного оператора *edge-aware* весовые функции конструируются строго через нормированные конечные разности. Асимптотическое поведение дискретных энергий исследуется с помощью аппарата сходимости.
Результаты исследования. Главным теоретическим результатом работы является математическое доказательство Γ-сходимости семейства дискретных анизотропных функционалов к нетривиальному непрерывному пределу в топологии $L^2(\Omega)$. Доказана равнокоэрцитивность дискретных энергий, гарантирующая сходимость последовательности почти минимизаторов к решению непрерывной задачи.
Обсуждение. Использование нормированных конечных разностей при построении весов восстанавливает размерную однородность и обеспечивает строгую масштабную инвариантность дискретизации нелокальных операторов.
Заключение. Предложенный метод успешно связывает непрерывные вариационные постановки с дискретными предиктивными моделями, предоставляя теоретически обоснованный и вычислительно эффективный (дополнительные расходы инференса составляют 17–18 %) инструмент с контролируемой погрешностью.

Ключевые слова: семантическая сегментация, распределение Дирихле, оценка неопределённости, калибровка вероятностей, анизотропная регуляризация, энергия Дирихле, Г-сходимость, равнокоэрцитивность, медицинские изображения

Финансирование. Работа выполнена при поддержке Севастопольского государственного университета (проект № 42–01–09/319/2025–1).

Для цитирования. Щетинин Е.Ю., Шевчук А.А. Семантическая сегментация с оценкой неопределённости на основе модели Дирихле и анизотропной регуляризации. *Computational Mathematics and Information Technologies*. 2026;10(1):7–20. https://doi.org/10.23947/2587-8999-2026-10-1-7-20

Introduction. Semantic segmentation of medical images is a key task in computer vision for diagnostics. However, modern deep neural networks often fail to provide reliable uncertainty estimates for their predictions [1, 2]. In clinical practice, this limits their application, as physicians must be able to identify areas where the model is inaccurate for additional verification.

Existing uncertainty estimation methods fall into two main categories:

1. Stochastic methods (MC-dropout [3], deep ensembles [4]), which require multiple forward passes during inference, making them unsuitable for real-time applications.

2. Deterministic methods for direct prediction of distribution parameters (Evidential Deep Learning [5], Prior Networks [6]), which operate in a single pass but often suffer from spatial inconsistency in uncertainty maps.

A number of studies investigate the reliability of uncertainty estimates in medical segmentation [7] and the application of Bayesian methods for lesion detection [8]. Issues of calibrating predictive distributions are considered in [9]. Recently, methods using the Dirichlet distribution for uncertainty estimation in medical segmentation have emerged, including approaches based on evidence theory [10] and adaptive methods [11]. Structural approaches to uncertainty [12] and methods robust to noisy labels [13] are also being developed.

The main objectives of this work are:

1. To formulate the problem as the minimization of a composite functional with a Dirichlet distribution and anisotropic spatial regularization.

2. To define edge-aware weights through normalized finite differences, ensuring a non-trivial continuous limit.

3. To prove the Γ-convergence of discrete anisotropic energies to a continuous limit, covering key technical steps (lower semicontinuity of weighted integrals and equicoercivity/compactness).

4. To provide comprehensive experimental validation with an extended statistical protocol, including appropriate effect size measures for paired data. The issues of robustness to out-of-distribution (OOD) data require separate investigation and are beyond the scope of this work.

Recall that the Dirichlet distribution is a probability distribution defined on the simplex; the Dirichlet energy is a functional of the form $\int_\Omega \|\nabla u(\xi)\|^2 \, d\xi$; equicoercivity refers to the coercivity of a family of functionals being uniform with respect to the discretization parameter.

Materials and Methods. Problem Statement and Mathematical Model.

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with a Lipschitz boundary, where $d \in \{2,3\}$.

Definition 1 (Function spaces). Let $L^p(\Omega;\mathbb{R}^r)$ be the Lebesgue space of measurable functions $u : \Omega \to \mathbb{R}^r$ with the norm

$$\| u \|_{L^p} = \left(\int_\Omega \| u(\xi) \|^p_{\mathbb{R}^r} \, d\xi \right)^{1/p}, \tag{1}$$

where $\|u(\xi)\|_{\mathbb{R}^r} = \left(\sum_{j=1}^{r} | u_j(\xi) |^2 \right)^{1/2}$ is the Euclidean norm in $\mathbb{R}^r$; $W^{1,2}(\Omega;\mathbb{R}^r)$ is the Sobolev space of functions with square-integrable weak first-order partial derivatives; $\Delta^{K-1} = \left\{ p \in \mathbb{R}^K_+ : \sum_{k=1}^{K} p_k = 1 \right\}$ is the standard $(K-1)$-dimensional simplex.

Notation for norms. For a matrix $A \in \mathbb{R}^{r\times d}$ (in particular, for the Jacobian matrix ∇_u) the Frobenius norm is used:

$$\| A \|_F^2 = \sum_{j=1}^{r} \sum_{k=1}^{d} |A_{jk}|^2 . \tag{2}$$

For a vector $v \in \mathbb{R}^r$ the Euclidean norm $\| v \|^2_{\mathbb{R}^r} = \sum_{j=1}^{r} | v_j |^2$ is used. When the type of norm is clear from the context, indices are omitted.

Domain Discretization.

Definition 2 (Uniform grid). Let $\Omega = (0,1)^d$, $d \in \{2,3\}$, $\xi = (\xi_1,...,\xi_d) \in \Omega$. For an integer $N \ge 2$ set $h := \frac{1}{N}$. We define a uniform Cartesian grid with step h as the set of nodes

$$\Omega_h := \Omega \cap h\mathbb{Z}^d = \left\{ x = (i_1 h, \ldots, i_d h) \in \Omega : i_k \in \mathbb{Z} \right\}. \tag{3}$$

The corresponding partition $\mathcal{T}_h$ consists of equal hyperrectangular cells $K_i := \prod_{k=1}^{d} \left[i_k h, (i_k+1)h \right]$, $i = (i_1, \ldots, i_d) \in Z_d$, such that $\bar{\Omega} = \bigcup_{\mathbf{i}} K_{\mathbf{i}}$ and the interior cells all have the same size $h \times \cdots \times h$. If necessary, one can distinguish between interior $\mathring{\Omega}_h := \{\xi_{\mathbf{i}} \in \Omega : 1 \le i_k \le N-1\}$ and boundary $\partial\Omega_h := \Omega_h \setminus \mathring{\Omega}_h$ nodes.

Assumption (G) (Grid regularity). A family of partitions $\{\mathcal{T}_h\}_{h\downarrow 0}$ is called shape-regular if there exists a constant $G_G > 0$ independent of h, such that for any cell $K \in \mathcal{T}_h$ we have $\frac{h_K}{\rho_K} \le C_G$, $h_K := \operatorname{diam}(K)$, $\rho_K := \sup\{r > 0 : \exists B_r(y) \subset K\}$, where ρ_K is the inradius of the cell K.

Remark for the uniform grid (3).

For $K_{\mathbf{i}} = [i_1 h, (i_1+1)h] \times \ldots \times [i_d h, (i_d+1)h]$ we have $h_K = \sqrt{d}\, h$, $\rho_K = \frac{h}{2}$, therefore $\frac{h_K}{\rho_K} = 2\sqrt{d}$, and Assumption (G) holds with $C_G = 2\sqrt{d}$ (in general, $C_G = 2\sqrt{d}$ for $d = 2$ and $C_G = 2\sqrt{3}$ for $d = 3$).

Definition 3 (Sets of nodes). The full set of interior nodes: $\mathring{\Omega}_h = \{x \in \Omega_h : \operatorname{dist}(x, \partial\Omega) > h\sqrt{d}\}$. Interior nodes with respect to the direction $k \in \{1,...,d\}$:

$$\Omega_h^{(k)} = \{x \in \Omega_h : x + he_k \in \Omega_h\}, \tag{4}$$

where e_k is the k-th canonical basis vector in $\mathbb{R}^d$.

Definition 4 (Spaces of grid functions). Scalar functions: $V_h = \{v_h : \Omega_h \to \mathbb{R}\}$. Vector functions: $V_h^r = \{u_h : \Omega_h \to \mathbb{R}^r\}$.

The space V_h is endowed with the discrete L^2 norm:

$$\| v_h \|_h^2 = h^d \sum_{x\in\Omega_h} |v_h(x)|^2 . \tag{5}$$

Similarly for V_h^r:

$$\| u_h \|_h^2 = h^d \sum_{x\in\Omega_h} \| u_h(x) \|^2_{\mathbb{R}^r} . \tag{6}$$

Discretization and Interpolation Operators.

Definition 5 (Restriction operator). For $u \in C^0(\bar{\Omega};\mathbb{R}^r)$ define the restriction operator as

$$\Pi_h : C^0(\bar{\Omega};\mathbb{R}^r) \to V_h^r : (\Pi_h u)(x) = u(x),\ x \in \Omega_h. \tag{7}$$

Definition 6 (Normalized finite difference). For $v_h \in V_h$ and $k \in \{1,...,d\}$:

$$D_k^+ v_h(x) = \frac{v_h(x+he_k) - v_h(x)}{h},\ x \in \Omega_h^{(k)}. \tag{8}$$

For $u_h \in V_h^r$ the definition is applied componentwise:

$$(D_k^+ u_h)_j(x) = D_k^+(u_h)_j(x),\ j = 1, \ldots, r. \tag{9}$$

Definition 7 (Multilinear interpolation). The operator $\mathcal{T}_h : V_h^r \to W^{1,2}(\Omega;\mathbb{R}^r)$ is the standard Q_1 interpolation (bilinear when $d = 2$ trilinear when $d = 3$ defined) by the conditions:

1. $(\mathcal{T}_h u_h)(x) = u_h(x)$ for all $x \in \Omega_h$.

2. On each cell C_x the function $\mathcal{T}_h u_h$ is a polynomial of degree at most one in each variable.

Definition 8 (Piecewise constant extension). For $v_h \in V_h$ define $v_h \in L^\infty(\Omega)$:

$$v_h(\xi) = v_h(x),\ \xi \in C_x = [x_1, x_1 + h) \times \ldots \times [x_d, x_d + h), \tag{10}$$

where $x \in \Omega_h$ is the lower left corner of the cell ξ. The operator u_h is defined similarly for $u_h \in V_h^r$.

Probabilistic Model Based on the Dirichlet Distribution.

Let $I : \Omega \to R^C$ be an input image (a continuous function or its discretization $I_h = \Pi_h I$), $y : \Omega_L \to \{1,...,K\}$ be the ground truth label at a subset of nodes $\Omega_L \subseteq \Omega_h$. At each node $x \in \Omega_h$ we model the vector of class probabilities as $p(x) \in \Delta^{K-1}$:

$$p(\mathbf{p}(x) \mid \boldsymbol{\alpha}(x)) = Dir(\mathbf{p}(x) \mid \boldsymbol{\alpha}(x)) = \frac{1}{B(\boldsymbol{\alpha}(x))} \prod_{k=1}^{K} p_k(x)^{\alpha_k(x)-1}, \tag{11}$$

where $\boldsymbol{\alpha}(x) = (\alpha_1(x), \ldots, \alpha_K(x)) \in \mathbb{R}^K_{>0}$ are concentration parameters, $B(\boldsymbol{\alpha}) = \prod_k \Gamma(\alpha_k) / \Gamma(\sum_k \alpha_k)$ being the multivariate beta function.

Predictive mean:

$$\mathbf{m}(x) = \mathbb{E}[\mathbf{p}(x) \mid \boldsymbol{\alpha}(x)] = \frac{\boldsymbol{\alpha}(x)}{S(x)},\ S(x) = \sum_{k=1}^{K} \alpha_k(x). \tag{12}$$

Uncertainty Map.

Lemma 1 (Mutual information as an Uncertainty Measure). For the hierarchical model $Y \mid \mathbf{p} \sim \mathrm{Cat}(\mathbf{p}),\ \mathbf{p} \mid \boldsymbol{\alpha} \sim \mathrm{Dir}(\boldsymbol{\alpha})$ the mutual information between the label and the probability vector $U(x) = I(Y;\ \mathbf{p} \mid \boldsymbol{\alpha})$ has the form:

$$U(x) = H[\mathbf{m}(x)] - \left(\psi(S(x)+1) - \sum_{k=1}^{K} \frac{\alpha_k(x)}{S(x)} \psi(\alpha_k(x)+1) \right), \tag{13}$$

where $H[\mathbf{m}] = -\sum_k m_k \log m_k$ is the Shannon entropy and, ψ is the digamma function.

In particular, the discrete regularizer Γ-converges to the continuous one:

$$\mathcal{J}^{\mathrm{reg}} = \beta_{\mathrm{KL}} R_{\mathrm{KL}} + \lambda_m R_m + \lambda_s R_s. \tag{14}$$

Proof. Follows from the expansion $I(Y;\mathbf{p}) = H(Y) - \mathbb{E}_{\mathbf{p}}[H(Y \mid \mathbf{p})]$ and the properties of the Dirichlet distribution [5, Appendix A].

Definition 9 (Edge-aware weights). For an image $I \in C^0(\bar{\Omega}; \mathbb{R}^C)$ and parameters $\eta > 0, \varepsilon > 0$ define the discrete weights:

$$w_h^k(x) = \exp\left(-\eta \, \| D_k^+ (\Pi_h I)(x) \|^2_{\mathbb{R}^C}\right) + \varepsilon,\ x \in \Omega_h^{(k)}. \tag{15}$$

Remark 1. The use of the normalized difference D_k^+ (division by h) is critical: it ensures convergence to a non-trivial limit function

$$w^k(\xi) = \exp\left(-\eta \, \| \partial_k I(\xi) \|^2_{\mathbb{R}^C}\right) + \varepsilon \tag{16}$$

as $h \to 0$ whereas unnormalized differences would lead to $w_h^k \to 1 + \varepsilon \equiv \mathrm{const}$ pointwise, trivializing the anisotropic properties.

Definition 10 (Composite functional). The full functional for optimization is:

$$\mathcal{J}_h[\boldsymbol{\alpha}_h] = \mathcal{L}_{data}[\boldsymbol{\alpha}_h] + \beta_{KL} \mathcal{R}_{KL}[\boldsymbol{\alpha}_h] + \lambda_m \mathcal{R}_m[\mathbf{m}_h] + \lambda_s \mathcal{R}_s[S_h], \tag{17}$$

where $\boldsymbol{\alpha}_h \in V_h^K, \mathbf{m}_h = \boldsymbol{\alpha}_h / S_h \in V_h^K,\ S_h = \sum_{k=1}^{K} (\boldsymbol{\alpha}_h)_k \in V_h$.

Data fidelity (expected log-loss):

$$\mathcal{L}_{data}[\boldsymbol{\alpha}_h] = h^d \sum_{x \in \Omega_L} \left(\psi(S_h(x)) - \psi((\boldsymbol{\alpha}_h)_{y(x)}(x)) \right). \tag{18}$$

KL-regularization towards a uniform prior:

$$\mathcal{R}_{KL}[\boldsymbol{\alpha}_h] = h^d \sum_{x \in \Omega_h} KL\left(Dir(\boldsymbol{\alpha}_h(x)) \,\|\, Dir(\mathbf{1}) \right). \tag{19}$$

Anisotropic spatial regularizers:

$$\mathcal{R}_m[\mathbf{m}_h] = h^d \sum_{k=1}^{d} \sum_{x \in \Omega_h^{(k)}} w_h^k(x) \, \| D_k^+ \mathbf{m}_h(x) \|^2_{\mathbb{R}^K}, \tag{20}$$

$$\mathcal{R}_s[S_h] = h^d \sum_{k=1}^{d} \sum_{x \in \Omega_h^{(k)}} w_h^k(x) \, | D_k^+ \log S_h(x) |^2. \tag{21}$$

Theoretical Justification. In this section, we fix a correct (scale-consistent) discretization of the anisotropic Dirichlet energy and formulate the Γ-convergence of the corresponding discrete functionals. Detailed proofs are standard and rely on classical results of Γ-convergence [14–16].

Continuous and discrete energies. Let $W^{1,2}(\Omega;\mathbb{R}^r)$ and $w^k \in L^\infty(\Omega)$, $k = 1,...,d$.

Continuous energy: $E[u]=\sum_{k=1}^{d}\int_{\Omega} w^k(\xi)\,\|\partial_k u(\xi)\|_{\mathbb{R}^r}^2\,d\xi.$

Discrete energy for $u_h \in V_h^r$ is defined via normalized finite differences $D_k^+ u_h(x)=(u_h(x+he_k)-u_h(x))/h$:

$$E_h[u_h]=h^d\sum_{k=1}^{d}\sum_{x\in\Omega_h^{(k)}} w_h^k(x)\,\|D_k^+u_h(x)\|_{\mathbb{R}^r}^2\,. \tag{22}$$

Conditions on the weights:

$(W1)$ $\varepsilon \le w_h^k(x)\le \overline{w}$; $(W2)$ the piecewise constant extension $w_h^k \to w^k$ in $L_1(\Omega)$.

Theorem 1 (Γ-convergence of energies). Let the weights satisfy conditions $(W1)$–$(W2)$. Then the discrete energies E_h Γ-converge to the continuous energy E in the $L^2(\Omega;\mathbb{R})$, topology, and the family E_h is equicoercive on the subspace of functions with fixed mean value (or under periodic boundary conditions). Consequently, every sequence of almost minimizers has a (possibly after extracting a subsequence) limit u which is a minimizer of E.

Proof.

1) *Equicoercivity*. Let $\{u_h\}\in V_h^r$ be a sequence $\sup_h E_h[u_h]<\infty$. From the lower bound $(W1)$ we obtain control over the discrete gradients:

$$\sum_{k=1}^{d}\|D_k^+u_h\|_{0,h}^2\le\frac{1}{\varepsilon}E_h[u_h]. \tag{23}$$

With the chosen calibration (fixed mean or periodic/fixing boundary conditions), the discrete Poincaré inequality provides control of the L^2 norm, and the sequence $\{u_h\}$ becomes compact in $L^2(\Omega;\mathbb{R}^r)$ after interpolation.

Passing to the multilinear interpolation $I_h u_h$, we use the standard stability estimate for interpolation:

$$\|\nabla(I_h u_h)\|_{L^2(\Omega)}\le C\,\|\nabla_h u_h\|_{0,h}\,. \tag{24}$$

Consequently, $\{I_h u_h\}$ is bounded in $H^1(\Omega;\mathbb{R}^r)$. By the Rellich–Kondrachov theorem, there exists a subsequence such that $I_h u_h \rightharpoonup u$ в $H^1(\Omega;\mathbb{R}^r)$ and $I_h u_h \to u$ в $L^2(\Omega;\mathbb{R}^r)$. This yields equicoercivity in the L^2 topology.

2) *Liminf inequality*. Let $u_h \to u$ в $L^2(\Omega;\mathbb{R}^r)$ and $\sup_h E_h[u_h]<\infty$. Then, as above, we may assume that $I_h u_h \rightharpoonup u$ in H^1 and $D_k^+u_h \rightharpoonup \partial_k u$ in L^2 (after extension to Ω). By $(W2)$ we have $w_h^k \to w^k$ in $L\infty$. For each direction k we use the weak lower semicontinuity of the quadratic form; we obtain

$$\int_{\Omega} w^k\left|\partial_k u\right|^2\le\liminf_{h\to0}\int_{\Omega} w^k\left|D_k^+u_h\right|^2. \tag{25}$$

The replacement of w_h^k by w^k is controlled by the estimate

$$\left|\int_{\Omega}\left(w_h^k-w^k\right)\left|D_k^+u_h\right|^2\right|\le\|w_h^k-w^k\|_{L^\infty(\Omega)}\|D_k^+u_h\|_{L^2(\Omega)}^2\to0. \tag{26}$$

Summing over $k = 1,\dots,d$, we obtain the liminf inequality $E_h[u_h] \ge E[u]$.

3) *Limsup inequality* (recovery sequence).

Let $u \in H^1(\Omega;\mathbb{R}^r)$. We construct a sequence $u_h\in\mathbb{R}^{r|\mathring{\Omega}_h|}$ as follows.

Case A (nodal discretization / restriction to nodes). Assume that $(u_h)_i := u(\xi_i)$, $\xi_i\in\mathring{\Omega}_h$, and consider piecewise linear (or piecewise constant) interpolation $\mathcal{I}_h u_h \in H^1(\Omega;\mathbb{R}^r)$.

Case B (cell projection / L^2 discretization). Assume that

$$(u_h)_i := \frac{1}{|K_i|}\int_{K_i}u(x)\,dx,\ u_h:=\Pi_h u, \tag{27}$$

where $K_i\in\mathcal{T}_h$ is the cell associated with the node i, and Π_h denotes the cell-averaging operator (L^2 projection onto the space of piecewise constant functions). In this case, $\mathcal{T}_h u_h$ is a piecewise constant function on $\mathcal{T}_h$.

In both cases (given the quasi-uniformity of the grid and standard properties of the Π_h, $\mathcal{T}_h$ projection), the approximation of the gradient holds: $D_k^+u_h \to \partial_k u$ in $L^2(\Omega;\mathbb{R}^r)$, $k = 1,...,d$, as well as $\mathcal{I}_h u_h \to u$ in $L^2(\Omega;\mathbb{R}^r)$. Furthermore, from the assumptions on the weights $(W2)$ and boundedness $(W1)$ we have $w_h^k \to w^k\,в\,L^\infty(\Omega)$, $0<\varepsilon\le w_h^k(x)\le M$. Consequently, using the convergence of the discrete gradients and the convergence of Riemann sums to integrals, we obtain $E_h[u_h] \to E[u]$. Thus, for a fixed smooth $u \in H^1(\Omega;\mathbb{R}^r)$ a recovery sequence u_h, has been constructed satisfying $\limsup_{h\to0}E_h[u_h]\le E[u]$.

Now let $u \in L^2(\Omega;\mathbb{R}^r)$ and $E[u] < 0$. Then $u \in H^1(\Omega;\mathbb{R}^r)$. Let us construct a sequence $u^{(m)}\in C^\infty(\Omega;\mathbb{R}^r)\cap H^1(\Omega;\mathbb{R}^r)$, such that $u^{(m)}\to u$ в $H^1(\Omega;\mathbb{R}^r)$, $E[u^{(m)}]\to E[u]$. For each m construct $u_h^{(m)}$ by one of the cases A/B above, so that $\lim_{h\to0}E_h\left[u_h^{(m)}\right]=E[u^{(m)}]$. Choosing a diagonal subsequence $h=h(m)\downarrow$ we obtain a sequence $u_{h(m)}:=u_{h(m)}^{(m)}$ for which $u_{h(m)}\to u$ в $L^2(\Omega;\mathbb{R}^r)$, $\limsup_{m\to\infty}E_{h(m)}[u_{h(m)}]\le E[u]$. This proves the limsup inequality for all u with finite limit energy.

Thus, $E_h \xrightarrow{\Gamma} E$ in the $L^2(\Omega;\mathbb{R}^r)$ topology, and the family $\{E_h\}$ is equicoercive by the fundamental theorem of Γ-convergence on the convergence of almost minimizers [14]. It follows that every sequence of almost minimizers $\{u_h\}$ has a convergent subsequence L^2 whose limit is a minimizer of the limit functional, and the minimum values $\inf E_h$ converge to $\inf E$.

Theorem 2 (Γ-convergence of regularizers). Let the discrete regularizers $R_m[m_h]$ and $R_s[S_h]$ be given by formulas (19)–(20), and let the weights satisfy $(W1)$–$(W2)$. Assume also that $S_h(x) \ge S_{\min} > 0$ (in implementation, this is ensured by the parametrization $\alpha_k(x) = 1 + \text{softplus}(s_k(x))$, so $\log S_h$ is well-defined). Then R_m and R_s Γ-converge in the L^2 topology to the functionals:

$$R_m(m) = \sum_{k=1}^{d} \int_\Omega w^k(\xi) \left\| \partial_k m(\xi) \right\|^2 d\xi,\ m \in H^1(\Omega;\mathbb{R}^K), \tag{28}$$

$$R_s(S) = \sum_{k=1}^{d} \int_\Omega w^k(\xi) \left| \partial_k \log S(\xi) \right|^2 d\xi,\ \log S \in H^1(\Omega). \tag{29}$$

Proof.

The point for R_m follows from Theorem 1, applied componentwise to the vector field m_h: the discrete functional (19) is a sum over directions of quadratic energies of the type E_h, and the Γ-limit is the sum of the corresponding integrals. For R_s we apply Theorem 1 to the scalar function $u_h = \log S_h$ (the condition $S_h \ge S_{\min}$ guarantees correctness). The Γ-convergence of the linear combination J_r^heg follows from the stability of Γ-convergence with respect to the addition of functionals and multiplication by positive constants.

Remark 1′ (elimination of the gradient kernel). For correct coercivity of the quadratic energies, we fix a gauge: either consider functions with zero mean, impose periodic boundary conditions, or fix the value at a single point. The choice does not affect the Γ-limit and is used only to exclude the addition of a constant.

Neural Network Architecture.

The base architecture used is U-Net [17] with a ResNet-34 [18] encoder pre-trained on ImageNet.

Output parametrization. The network predicts logits $\mathbf{s}(x) \in R^K$ each pixel x. The Dirichlet parameters are computed as

$$\alpha_k(x) = 1 + \text{softplus}(s_k(x)) = 1 + \log(1 + \exp(s_k(x))),\ k = 1,\ldots,K. \tag{30}$$

This guarantees $\alpha_k(x) > 1$ which corresponds to a unimodal Dirichlet distribution.

Predictive probabilities:

$$\mathbf{m}(x) = \frac{\boldsymbol{\alpha}(x)}{S(x)},\ S(x) = \sum_{k=1}^{K} \alpha_k(x). \tag{31}$$

Edge-aware weights are computed once before training for each image I:

$$w_h^k(x) = \exp\left(-\eta \left\| \frac{I(x + he_k) - I(x)}{h} \right\|^2 \right) + \varepsilon. \tag{32}$$

In practice, for images with discrete pixels $h = 1$ (one pixel), and the formula simplifies to:

$$w_{i,j}^k = \exp\left(-\eta \, \| I_{i+\delta_k, j+\delta_l} - I_{i,j} \|^2 \right) + \varepsilon, \tag{33}$$

where $(\delta_k, \delta_l) \in ((1,0), (0,1))$ for $d = 2$.

Remark 2. When changing the image resolution (e. g., when working with pyramids), the parameter η does not require retuning due to the normalization by h.

Optimization Algorithm. Model Training.

Input: Training dataset $((I^{(n)}, y^{(n)}))_{n=1}^N$ hyperparameters β_{KL}, λ_m, λ_s, η, ε.

Output: Network parameters θ*.

Algorithm steps:

1. Precomputation of weights: for each $I^{(n)}$ compute w_h^k.
2. Initialization: $\theta \to \theta_0$ (pre-trained weights).
3. Loop over epochs: For $t = 1,\ldots,T$:

For each mini-batch $\mathcal{B}$:

Forward pass: $\boldsymbol{\alpha}^{(n)} = f_\theta(I^{(n)})$ for $n \in \mathcal{B}$.

Compute the loss functional (16).

Backward pass and optimizer step (Adam [19]).

4. Return: θ* = θ.

The additional cost of the proposed method (+18%) (Table 1) is due to:

− computing K parameters α_k instead of K logits (negligible);

− precomputing the edge-aware weights w_h^k (once per image).

Table 1

Time costs of the method (NVIDIA A100, image batch size = 1, inference only)

Method	Time (ms)	Relative
Baseline (CE)	22±1	100%
Proposed method	26±1	118%
MC-dropout (T = 30)	660±15	3000%
Deep Ensembles (M = 5)	110±5	500%

Hyperparameter selection protocol (Table 2). Hyperparameters were tuned using 5-fold cross-validation exclusively on the *train+val* set. The test set was not used at any stage of model or parameter selection. After fixing the hyperparameters, the following procedure was performed:

1. Final training on the *train* set.
2. Early stopping based on the *val* set.
3. Evaluation on the *test* set.

Table 2

Dataset characteristics

Dataset	Modality	*Train*	*Val*	*Test*	Классы
ACDC [20]	Cardiac MRI	70	10	20	4
Synapse [21]	Abdominal CT	18	6	6	9
CHAOS [22]	Liver CT/MRI	24	8	8	4

Note (patient-wise *split*). In all datasets, the *train/val/test* split was performed at the patient (volume) level, rather than on individual 2D slices, which prevents information leakage between subsets. Results are averaged over 10 independent runs with different random seeds. All methods were trained under identical conditions: fixed data splits and the same initial weights for comparable architectures.

Quality Metrics.

Segmentation:

$$\text{Dice} = 2\,|A \cap B|/(|A| + |B|),\ \text{IoU} = |A \cap B|/|A \cup B|. \quad (34)$$

Calibration: ECE (*Expected Calibration Error*): 15 uniform bins by confidence, micro-averaging over pixels of all classes (including background):

$$\text{ECE} = \sum_{b=1}^{15} \frac{n_b}{N}\,|\,\text{acc}(b) - \text{conf}(b)\,|, \quad (35)$$

where b is the number of pixels in bin, is the fraction of correct predictions, the average confidence. NLL (*Negative Log-Likelihood*): average over pixels $-\log m_{y(x)}(x)$.

Error Detection. *AUROC*: area under the ROC curve for the binary task “correct/incorrect prediction” using uncertainty $U(x)$ as the *score*. *AURC* (*Area Under Risk-Coverage Curve*): an evaluation metric for selective segmentation [23].

Statistical Analysis. Effect size for paired data. Since comparisons are conducted on the same test examples (paired design), *Cohen’s* is used $d_z : d_z = \frac{\bar{D}}{s_D}$, $D_i = X_i^{\text{our}} - X_i^{\text{baseline}}$ where $\bar{D}$ is the mean of the differences and s_D is the standard deviation of the differences.

Normality check. The Shapiro-Wilk test was applied to the differences for each metric. The Wilcoxon signed-rank test was used; otherwise, the paired t-test was employed.

Correction for multiple comparisons. The Holm-Bonferroni correction was applied within each family of hypotheses (comparisons with the baseline for a single dataset). For comparisons between datasets, FDR control according to Benjamini-Hochberg was used.

Confidence intervals. 95% BCa bootstrap intervals (10 000 replicates) for all metrics.

Compared Methods:

1. CE (*baseline*): cross-entropy loss, U-Net + ResNet-34.
2. *MC-dropout* [3]: *dropout* p = 0.5 at inference, T = 30 forward passes.
3. *Deep Ensembles* [4]: M = 5 independently trained models.
4. *Evidential U-Net* [10]: Dirichlet distribution without spatial regularization.
5. *UDEL* [11]: adaptive uncertainty estimation.

6. Proposed method: full functional (16).
7. *nnU-Net* [24]: external benchmark (self-configuring pipeline, not strictly comparable).

Research Results

Table 3

Segmentation Quality and Calibration on ACDC (mean ± std, 10 runs)

Method	*Dice*	*ECE*	*NLL*
CE	0.891 ± 0.011	0.078 ± 0.006	0.351 ± 0.018
MC-dropout	0.897 ± 0.010	0.032 ± 0.004	0.315 ± 0.014
Deep Ensembles	0.908 ± 0.008	0.024 ± 0.003	0.301 ± 0.012
Evidential U-Net	0.909 ± 0.008	0.023 ± 0.003	0.303 ± 0.013
UDEL	0.907 ± 0.009	0.025 ± 0.004	0.306 ± 0.014
Proposed method	**0.912 ± 0.008**	**0.021 ± 0.003**	**0.298 ± 0.012**
nnU-Net	0.915 ± 0.007	0.023 ± 0.003	0.295 ± 0.011

Table 4

Error Detection and Selective Segmentation (ACDC)

Method	*AUROC*	95% *CI*	*AURC*	95% *CI*
CE (энтропия)	0.812 ± 0.015	[0.798, 0.826]	0.178 ± 0.012	[0.166, 0.190]
MC-dropout	0.847 ± 0.013	[0.835, 0.859]	0.156 ± 0.010	[0.146, 0.166]
Deep Ensembles	0.875 ± 0.011	[0.865, 0.885]	0.132 ± 0.008	[0.124, 0.140]
Evidential U-Net	0.8885 ± 0.010	[0.876, 0.894]	0.129 ± 0.007	[0.122, 0.136]
UDEL	0.882 ± 0.011	[0.872, 0.892]	0.131 ± 0.008	[0.123, 0.139]
Proposed method	**0.891 ± 0.009**	[0.883, 0.899]	**0.124 ± 0.006**	[0.118, 0.130]

The Shapiro-Wilk test for the differences: $p > 0.15$ for all metrics, normality is not rejected (Table 5).

Table 5

Statistical Significance (Proposed method vs. CE baseline, ACDC)

Metric	Δ (Proposed — *CE*)	95% *BCa CI*	*Cohen's*	*p(Holm)*	Test
Dice	+0.021	[0.014, 0.028]	2.1	< 0.001	*t*-test
ECE	−0.057	[−0.063, −0.051]	2.3	< 0.001	*t*-test
NLL	−0.053	[−0.068, −0.038]	1.9	< 0.001	*t*-test
AUROC	+0.079	[0.065, 0.093]	2.4	< 0.001	*t*-test

Results on the Synapse and CHAOS datasets are presented in Table 6.

Empirical Verification of Theoretical Results.

Experiment 1. Convergence order of the discrete energy.

Test profile: $u(\xi) = \sin(\pi\xi_1)\cos(\pi\xi_2)$, $w^k(\xi) = 1 + 0,5\cos(2\pi\xi_k)$, $\Omega = [0,1]^2$. Analytical value: $E[u] = 4.9348$.

Table 6

Results on Synapse and CHAOS

Dataset	Method	*Dice*	*ECE*	*AUROC*
Synapse	*CE*	0.762 ± 0.014	0.081 ± 0.007	0.847 ± 0.014
Synapse	*Deep Ensembles*	0.785 ± 0.010	0.027 ± 0.004	0.887 ± 0.010
Synapse	Proposed method	**0.789 ± 0.011**	**0.024 ± 0.004**	**0.895 ± 0.008**
CHAOS	*CE*	0.883 ± 0.012	−	0.831 ± 0.016
CHAOS	*Deep Ensembles*	0.901 ± 0.009	0.039 ± 0.005	0.872 ± 0.012
CHAOS	Proposed method	**0.906 ± 0.009**	**0.036 ± 0.004**	**0.883 ± 0.010**

Table 7

Convergence of the discrete energy (single test profile)

h	E_h	$\lvert E_h - E \rvert$	Order
1/8	4.8721	0.0627	–
1/16	4.9035	0.0313	1.00
1/32	4.9191	0.0157	0.99
1/64	4.9270	0.0078	1.01
1/128	4.9309	0.0039	1.00

To confirm the universality of the order, experiments were conducted on additional profiles (Table 8): $\xi = (\xi_1, \xi_2) \in (0,1)^2$, $k \in \{1,2\}$. For B: $u(\xi) = \xi_1 + \xi_2$, $w^k(\xi) \equiv 1$. For D: $u(\xi) = \exp(-(\xi_1^2 + \xi_2^2))$, $w^k(\xi) = \exp(-\xi_k^2)$.

Table 8

Convergence order for various test functions

Profile	$u(\xi)$	$w^k(\xi)$	Order (h: 1/16 → 1/128)
A	$\sin(\pi\xi_1)\cos(\pi\xi_2)$	$1 + 0{,}5\cos(2\pi\xi_k)$	1.00 ± 0.01
B	$\xi_1 + \xi_2$	1	1.01 ± 0.01
C	$\xi_1^2(1-\xi_1)^2\xi_2^2(1-\xi_2)^2$	$1 + \xi_k$	0.99 ± 0.02
D	$exp(-(\xi_1^2 + \xi_2^2))$	$exp(-\xi_k^2)$	0.99 ± 0.02

All profiles demonstrate first-order convergence, consistent with Theorem 1.

Ablation Analysis.

Table 9

Impact of functional components (ACDC, 10 runs)

Configuration	*Dice*	*ECE*	*NLL*	*AURC*
Full model	**0.912**	**0.021**	**0.298**	**0.124**
Without $\mathcal{R}_{\mathrm{KL}}$	0.897	0.052	0.325	0.151
Without $\mathcal{R}_m$	0.899	0.048	0.320	0.145
Without $\mathcal{R}_s$	0.905	0.035	0.310	0.136
Unnormalized differences in weights	0.908	0.028	0.305	0.131
CE baseline	0.891	0.078	0.351	0.178

− All functional components contribute to the improvement.
− The regularization $\mathcal{R}_{\mathrm{KL}}$ has the greatest effect on calibration.
− Normalized differences in weights improve ECE by 25% compared to unnormalized ones.

Reproducibility.

• Data and splitting: fixed patient-wise splits; unified preprocessing and postprocessing for all methods.
• Training: Adam optimizer; mixed precision (FP16/AMP); 10 different initializations (seeds).
• Metrics: Dice, Jaccard, ECE; for error detection — AUROC and AURC based on the uncertainty map; statistics — paired Wilcoxon or t-test (depending on normality) with Holm-Bonferroni correction.
• Hardware/software environment: NVIDIA A100 GPU; PyTorch; code and configurations available upon request / after acceptance.

Hyperparameter Sensitivity (ACDC)

Tables 11–12 show mean ± std over 10 initializations; one parameter was varied while others were taken from Table 10. Here λ denotes the joint value of $\lambda = \lambda_m = \lambda_s$.

Table 10

Optimal hyperparameters (tuned on validation set)

Parameter	Description	*ACDC*	*Synapse*	*CHAOS*
β_{KL}	Weight of KL regularization	0.100	0.100	0.080
λ_m	Weight of *m* regularization	0.010	0.015	0.012
λ_s	Weight of *S* regularization	0.010	0.015	0.012
η	Sensitivity weight parameter	5.000	4.500	5.500
ε	Lower bound for weights	1×10^{-3}	1×10^{-3}	1×10^{-3}
lr	Learning rate	1×10^{-4}	1×10^{-4}	1×10^{-4}
batch size	Mini-batch size	8	8	8
epochs	Number of epochs	100	150	100

Table 11

Effect of parameter η (ACDC)

η	*Dice*	*ECE*	*AUROC*
1.0	0.905 ± 0.009	0.028 ± 0.004	0.879 ± 0.011
2.5	0.908 ± 0.008	0.024 ± 0.003	0.885 ± 0.010
5.0	0.912 ± 0.008	0.021 ± 0.003	0.891 ± 0.009
10.0	0.910 ± 0.008	0.022 ± 0.003	0.888 ± 0.010
20.0	0.907 ± 0.009	0.025 ± 0.004	0.882 ± 0.011

Table 12

Effect of parameter λ (ACDC)

λ	*Dice*	*ECE*	*AUROC*
0.001	0.903 ± 0.010	0.031 ± 0.004	0.875 ± 0.012
0.005	0.908 ± 0.009	0,025 ± 0.003	0.884 ± 0.010
0.010	0.912 ± 0.008	0,021 ± 0.003	0.891 ± 0.009
0.020	0.909 ± 0.008	0,023 ± 0.003	0.887 ± 0.010
0.050	0.901 ± 0.011	0,027 ± 0.004	0.878 ± 0.012

Ablation of Weight Formulation (ACDC).

Comparisons were made between: (*a*) normalized differences with division by grid step *h*; (*b*) unnormalized differences (without division by *h*); (*c*) isotropic weights $w = \varepsilon + \exp(-\eta\cdot\|\nabla_h I\|^2)$; (d) constant weights $w \equiv 1$ (Table 13).

Table 13

Effect of weight formulation (ACDC)

Weight Formulation	*Dice*	*ECE*	*AUROC*
Normalized differences (proposed method)	0.912	0.021	0.891
Unnormalized differences	0.908	0.028	0.883
Isotropic weights $w = \varepsilon + \exp(-\eta\cdot\|\nabla_h I\|^2)$	0.909	0.024	0.886
Constant weights $w \equiv 1$	0.902	0.035	0.871

Metrics AUROC and AURC.

Let $\hat{y}(x)$ be the predicted class, $y(x)$ the ground truth label, $e(x) = 1\hat{y}(x) \neq y(x)$ the error indicator, and $U(x) \geq 0$ the scalar uncertainty. For error detection, we use $U(x)$ as the score:

$$e(x) = 1\{\hat{y}(x) \neq y(x)\}. \tag{36}$$

Define coverage and risk at threshold τ based on $U(x)$:

$$\mathrm{cov}(\tau) = \mathbb{P}(U(x) \leq \tau),\ \mathrm{risk}(\tau) = \mathbb{P}(e(x) = 1\ |U(x) \leq \tau). \tag{37}$$

AUROC is the area under the ROC curve of the binary classifier $e(x)$ using score $U(x)$; *AURC* is the area under the risk-coverage curve: $AURC = \int_0^1 risk(cov)\,d(cov)$, estimated discretely by ranking $U(x)$.

Discrete Grid and Difference Operators.

A uniform grid $\Omega_h \subset \Omega$ is defined with step $h = 1/N$. For direction $k = 1,\ldots,d$ we use the normalized forward difference and discrete inner product defined by formulas [26]:

$$D_{k,h}v_h(x) = \frac{v_h(x+he_k) - v_h(x)}{h}, \tag{38}$$

$$\langle a,b\rangle_h = h^d \sum_{x\in\Omega_h} a(x)b(x), \tag{39}$$

$$\Omega_h = \{(i_1h,\ldots,i_dh) : i_j = 0,1,\ldots,N\} \cap \Omega. \tag{40}$$

Properties of the Dirichlet distribution and its associated moments/entropies are described in the monograph [27].

Remark: We employ a regularizing formulation consistent with classical approaches to ill-posed problems [28].

Outline of the Γ-convergence proof.

The proof of Γ-convergence consists of three standard steps: equicoercivity (compactness) of the family E_h; the liminf estimate for an arbitrary convergent sequence $u_h \to u$ in $L^2(\Omega)$; and the construction of a recovery sequence for each $u \in H^1(\Omega)$. The classical Γ-convergence theorem on the convergence of almost minimizers is then applied.

Lemma 1. Let w^n, $w \in L\infty(\Omega)$ in $L^2(\Omega)$, and $0 < \varepsilon \le w_n$, $w \le M$, $w_n \to w$ in $L^2(\Omega)$. Then the following inequality holds:

$$\int_\Omega w(x)|f(x)|^2\,dx \le \liminf_{\{n\to\infty\}} \int_\Omega w_n(x)|f_n(x)|^2\,dx. \tag{41}$$

Proof. For $M_0 > 0$ we decompose the squared modulus into a bounded part and a tail part:

$$|f_n|^2 = \min(|f_n|^2, M_0) + (|f_n|^2 - M_0)_+. \tag{42}$$

The bounded part is controlled by the convergence of the weights, while the tail is controlled by the energy estimate:

$$\int_\Omega (w_n - w)\min(|f_n|^2, M_0) \le M_0 \| w_n - w \|_{L^1(\Omega)} \to 0. \tag{43}$$

For the tail, we have:

$$\int_\Omega (|f_n|^2 - M_0)_+ \le \frac{1}{\varepsilon M_0}\int_\Omega w_n|f_n|^2. \tag{44}$$

Choosing $M_0 \to \infty$ and using the weak lower semicontinuity of the quadratic form $f \mapsto \int w|f|^2$ for a fixed w, we obtain the desired statement.

Lemma 2. Let $u_h \subset V_h$ and $\sup_h E_h(u_h) < \infty$. Under a gauge that eliminates the kernel of the gradient (zero mean or periodic/fixing boundary conditions; see Remark 1′), there exists a subsequence $h_j \to 0$ and a function $u \in H^1(\Omega)$ such that $\Pi_{hj}\, u_{hj} \to u$ в $L^2(\Omega)$ in $L^2(\Omega)$, and the discrete gradients $D_k^+ u_{hj} \rightharpoonup \partial_k u$ in $L^2(\Omega)$ for $k = 1,\ldots,d$.

Proof. From the lower bound $\varepsilon \le w_h^k$ we obtain an estimate for the discrete gradients:

$$\| D^+ u_h \|^2_{L^2(\Omega)} \le \frac{1}{\varepsilon} E_h(u_h). \tag{45}$$

The discrete Poincaré inequality, under the chosen gauge, gives a uniform bound for the L^2 norm; subsequently, compactness and limit identification follow from standard results of finite-difference/finite-element approximation [16].

Linking and Corollary on Almost Minimizers.

Liminf inequality.

Let $u_h \to u$ in $L^2(\Omega;\mathbb{R}^r)$ and $\sup_h E_h[u_h] < \infty$. Then from equicoercivity (Lemma 2), it follows that there exists a subsequence such that $D^+u_h \rightharpoonup \nabla u$ в $L^2(\Omega; R^{r\times d})$, $u \in H^1(\Omega; \mathbb{R}^r)$. Let $u_h \to u$ in $L^2(\Omega)$. Furthermore, from ($W2$) and the construction of the weights (see Remark 1), we have $w_h^k \to w^k$ in $L^\infty(\Omega)$ ($k = 1,\ldots,d$), $0 < \varepsilon \le w_h^k(x) \le M$. Applying Lemma 1 to the sequence $f_h : D_k^+ u_h$ and weights w_h^k we obtain for each k: $\int_\Omega w^k |\partial_k u|^2\,dx \le \liminf_{h\to 0} \int_\Omega w_h^k |D_k^+ u_h|^2\,dx$. Summing over $k = 1,\ldots,d$, we conclude that $E[u] \le \liminf_{h\to 0} E_h[u_h]$, thus proving the liminf inequality of Γ-convergence.

Remark on the convergence norm for weights. For the liminf step, it is sufficient to have $w_h^k \to w^k$ in $L^2(\Omega)$ given the uniform boundedness of w_h^k in $L^\infty(\Omega)$ and the lower bound $w_h^k \ge \varepsilon > 0$. The authors employ the stronger convergence in $L^\infty(\Omega)$ which is natural for weights generated by a smoothed image.

Limsup inequality (recovery sequence). Let first $u \in H^1(\Omega;\mathbb{R}^r)$. We define the recovery sequence via a standard grid interpolation operator: $u_h := \mathcal{I}_h u$ where $\mathcal{I}_h$ is a chosen discretization (piecewise linear nodal interpolation or piecewise constant cell-based interpolation). Then $\mathcal{I}_h u \to u$ в $L^2(\Omega;\mathbb{R}^r)$, $D^+_k(\mathcal{I}_h u) \to \partial_k u$ в $L^2(\Omega;\mathbb{R}^r)$, $k = 1,\ldots,d$. Using the boundedness of w_h^k and the convergence $w_h^k \to w^k$ ($W2$), as well as the convergence of discrete sums to integrals, we obtain $\lim_{n\to 0} E_h[u_h] = E[u]$, and consequently, $\limsup_{h\to 0} E_h[u_h] \le E[u]$. For an arbitrary $u \in L^2(\Omega;\mathbb{R}^r)$ with $E[u] < 0$ we have a sequence $u \in H^1(\Omega; \mathbb{R}^r)$.

Thus, the limsup inequality is proven for all u with finite limit energy. This completes the proof of the Γ-convergence of the energies: $E_h \xrightarrow{\Gamma} E$ in $L^2(\Omega;\mathbb{R}^r)$.

Theorem 2 follows from Theorem 1 by applying Γ-convergence to each term of the regularizer and the standard stability properties of Γ-convergence: closure under finite sums of Γ-converging functionals and under the addition of continuous (in L^2) lower-order terms. Equicoercivity (Lemma 2) and Γ-convergence imply the standard result on the convergence of almost minimizers [14]: if $\delta_h \downarrow 0$ and $u_h - \delta_h$ are almost minimizers, then there exists a subsequence and $u^* \in$ Armin E such that $u_h \to u^*$ in $L^2(\Omega;\mathbb{R}^r)$ and $\lim_{n\to 0} \inf E_h = \inf E$, $\lim_{n\to 0} E_h[u_h] = E[u^*] = \inf E$.

Discussion

Main Results. The proposed method demonstrates:

1. Segmentation accuracy at the level of state-of-the-art methods: Dice = 0.912 on ACDC (Table 3), comparable to Deep Ensembles (0.908) and nnU-Net (0.915).

2. A significant improvement in calibration: ECE is reduced from 0.078 to 0.021 (a 73% improvement relative to the baseline). This means that the predicted class probabilities more accurately reflect the true frequencies.

3. Superior error detection performance: AUROC = 0.891 surpasses all compared methods (Table 4). The uncertainty maps effectively identify pixels with erroneous predictions.

4. Computational efficiency: 18% additional computational cost compared to 3000% for MC-dropout and 500% for Deep Ensembles.

Theoretical Justification. The proof of Γ-convergence (Theorem 2) ensures:

− Correctness of discretization: the discrete functional approximates the continuous one with controlled error.

− Convergence of minimizers: solutions of the discrete problems converge to the solution of the continuous problem.

− Justification of scale invariance: normalization of differences by h guarantees a non-trivial limit for the weights.

Comparison with Existing Approaches.

• *Evidential Deep Learning* [5]: utilizes the Dirichlet distribution but without spatial regularization. The proposed method adds anisotropic regularizers, improving the spatial consistency of uncertainty maps.

• *Anisotropic Diffusion* [25]: a classical approach to edge-preserving smoothing. This work provides, for the first time, a rigorous justification of Γ-convergence for edge-aware weights defined via discrete differences.

• *Deep Ensembles* [4]: provide good calibration at the cost of a multiple-fold increase in computational expense. The method proposed by the authors achieves comparable quality in a single forward pass.

Limitations.

1. 2D formulation. Experiments were conducted on 2D slices. Extension to full 3D volumes requires architecture adaptation and increased memory.

2. *In-distribution validation*. All experiments were performed on data from the same distribution. The behavior of the method on out-of-distribution (OOD) data (other modalities, pathologies, artifacts) requires separate investigation.

3. Condition $I \in W^{1,\infty}$. The theoretical results require boundedness of the image gradient. In practice, this is ensured by standard preprocessing (normalization, clipping).

4. Hyperparameter selection. The parameters η, ε, λ_m, λ_s are tuned on the validation set. Automatic selection methods (e. g., Bayesian optimization) could potentially improve the results.

Conclusion. This paper proposes a computational method for semantic segmentation with uncertainty estimation, based on the Dirichlet distribution and anisotropic spatial regularization. The ΓΓ-convergence of the discrete anisotropic energies to a continuous limit is proved; key technical lemmas on lower semicontinuity and equicoercivity are presented. A statistically significant improvement in calibration (Cohen's $d_z > 2, p < 0.001$ is achieved with minimal computational overhead (18%). The use of normalized finite differences in defining edge-aware weights ensures convergence to a non-trivial limit weight function.

The method demonstrated improvement on three different medical datasets (ACDC, Synapse, CHAOS) with various modalities (MRI, CT) and anatomical structures.

Directions for future research:

1. Extension to 3D. Generalizing the method to three-dimensional medical images (CT, MRI volumes) with a corresponding adaptation of the anisotropic regularization.

2. Analysis of OOD robustness. Investigating the method's behavior on out-of-distribution data, including other modalities, pathologies, and image artifacts.

3. Automatic hyperparameter selection. Developing methods for adaptive selection of $\eta, \lambda_m, \lambda_s$ based on input image properties.

4. Generalization of theoretical results. Extending the Γ-convergence analysis to the case of discontinuous weights $w^k \in BV(\Omega)$ which would allow consideration of images with sharp boundaries.

References

1. Begoli E., Bhattacharya T., Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat Mach Intell*. 2019;1(1):20–23. https://doi.org/10.1038/s42256-018-0004-1

2. Abdar M., Pourpanah F., Hussain S., et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf Fusion.* 2021; 76:243–297. https://doi.org/10.1016/j.inffus.2021.05.008

3. Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: *Proc. ICML*. New York: PMLR; 2016. P. 1050–1059.

4. Lakshminarayanan B., Pritzel A., Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. In: *Advances in Neural Information Processing Systems*. 2017;30:6402–6413.

5. Sensoy M., Kaplan L., Kandemir M. Evidential deep learning to quantify classification uncertainty. In: *Advances in Neural Information Processing Systems*. 2018;31:3183–3193.

6. Malinin A., Gales M. Predictive uncertainty estimation via prior networks. In: *Advances in Neural Information Processing Systems*. 2018;31:7047–7058.

7. Jungo A., Reyes M. Assessing reliability and challenges of uncertainty estimations for medical image segmentation. In: *MICCAI 2019*. LNCS, vol. 11765. Cham: Springer; 2019. P. 48–56. https://doi.org/10.1007/978-3-030-32245-8_6

8. Nair T., Precup D., Arnold D.L., Arbel T. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Med Image Anal.* 2020;59:101557. https://doi.org/10.1016/j.media.2019.101557

9. Mehrtash A., Wells W.M., Tempany C.M., Abolmaesumi P., Kapur T. Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE Trans Med Imaging*. 2020;39(12):3868–3878. https://doi.org/10.1109/TMI.2020.3006437

10. Li H., Nan Y., Del Ser J., Yang G. Region-based evidential deep learning to quantify uncertainty and improve robustness of brain tumor segmentation. *Neural Comput Appl*. 2023;35:22071–22085. https://doi.org/10.1007/s00521-022-08016-4

11. UDEL: Rethinking uncertainty dynamic estimation learning for ambiguous medical image segmentation. *Digit Signal Process*. 2025;169:105723. https://doi.org/10.1016/j.dsp.2025.105723

12. Yang B., Zhang X., Zhang H., et al. Structural uncertainty estimation for medical image segmentation. Med Image Anal. 2025;103:103602. https://doi.org/10.1016/j.media.2025.103602

13. Han K., Wang S., Chen J., et al., Region uncertainty estimation for medical image segmentation with noisy labels. IEEE Trans Med Imaging. 2025;44(12):5197–5207. https://doi.org/10.1109/TMI.2025.3589058

14. Dal Maso G. *An Introduction to Γ-Convergence*. Boston: Birkhäuser; 1993. https://doi.org/10.1007/978-1-4612-0327-8

15. Braides A. *Γ-Convergence for Beginners*. Oxford: Oxford University Press; 2002. https://doi.org/10.1093/acprof:oso/9780198507840.001.0001

16. Ciarlet P.G. *The Finite Element Method for Elliptic Problems*. Philadelphia: SIAM; 2002. https://doi.org/10.1137/1.9780898719208

17. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *MICCAI 2015*. LNCS, vol. 9351. Cham: Springer; 2015. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

18. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. In: *Proc. CVPR*. 2016. P. 770–778. https://doi.org/10.1109/CVPR.2016.90

19. Kingma D.P., Ba J. Adam: a method for stochastic optimization. In: *Proc. ICLR*. 2015. arXiv:1412.6980.

20. Bernard O., Lalande A., Zotti C., et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis. *IEEE Trans Med Imaging*. 2018;37(11):2514–2525. https://doi.org/10.1109/TMI.2018.2837502

21. Landman B.A., Xu Z., Iglesias J.E., et al. MICCAI Multi-Atlas Labeling Beyond the Cranial Vault — Workshop and Challenge. https://doi.org/10.7303/syn3193805

22. Kavur A.E., Gezer N.S., Bariş M., et al. CHAOS Challenge — combined (CT-MR) healthy abdominal organ segmentation. *Med Image Anal.* 2021;69:101950. https://doi.org/10.1016/j.media.2020.101950

23. Geifman Y., El-Yaniv R. Selective classification for deep neural networks. In: *Advances in Neural Information Processing Systems*. 2017;30:4878–4887.

24. Isensee F., Jaeger P.F., Kohl S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203–211. https://doi.org/10.1038/s41592-020-01008-z

25. Perona P., Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell. 1990;12(7):629–639. https://doi.org/10.1109/34.56205

26. Bakhvalov N.S., Zhidkov N.P., Kobelkov G.M. *Numerical methods*. Moscow: BINOM; 2012. 636 p.

27. Johnson N.L., Kotz S., Balakrishnan N. *Continuous Univariate Distributions*. Vol. 2. 2nd ed. New York: Wiley; 1995. (In Russ.)

28. Tikhonov A.N., Arsenin V.Ya. *Methods for solving ill-posed problems.* Moscow: Nauka; 1979. 288 p. (In Russ.)

29. Samarsky A.A. *The theory of difference schemes*. Moscow: Nauka; 1989. 616 p. (In Russ.)

30. Chen T., Xu B., Zhang C., Guestrin C. Training deep nets with sublinear memory cost. *arXiv:1604.06174*. 2016.

31. Micikevicius P., Narang S., Alben J. et al. Mixed precision training. In: *Proc. ICLR*. 2018. arXiv:1710.03740.

About the Authors:

Evgeny Yu. Shchetinin, Dr. Sci. (Phys.-Math.), Professor, Department of Information Technologies, Sevastopol State University (33, Universitetskaya St., Sevastopol, 299053, Russian Federation), ORCID, SPIN-code, riviera-molto@mail.ru

Andrey A. Shevchuk, PhD Student, Department of Information Technologies, Sevastopol State University (33, Universitetskaya St., Sevastopol, 299053, Russian Federation), ORCID, andreiluck11@yandex.ru

Contributions of the authors:

E.Yu. Shchetinin: general scientific supervision; problem statement; formulation of research ideas, objectives and goals; methodology development; derivation of calculation formulas and numerical experiment.

A.A. Shevchuk: formulation of achieved results and description of their significance; manuscript preparation.

Conflict of Interest Statement: **the authors declare no conflict of interest.**

All authors have read and approved the final manuscript.

Об авторах:

Евгений Юрьевич Щетинин, доктор физико-математических наук, профессор кафедры информационных технологий Севастопольского государственного университета (Российская Федерация, 299053, г. Севастополь, ул. Университетская, 33), ORCID, SPIN-код, riviera-molto@mail.ru

Андрей Андреевич Шевчук, аспирант кафедры информационных технологий Севастопольского государственного университета (Российская Федерация, 299053, г. Севастополь, ул. Университетская, 33), ORCID, andreiluck11@yandex.ru

Заявленный вклад авторов:

Е.Ю. Щетинин: общее научное руководство; постановка задачи; формулировка идей исследования, целей и задач, разработка методологии, получение расчетных формул и проведение численного эксперимента.

А.А. Шевчук: формулировка достигнутых результатов и описание их значимости, оформление материала статьи.

Конфликт интересов: **авторы заявляют об отсутствии конфликта интересов.**

Все авторы прочитали и одобрили окончательный вариант рукописи.

Received / Поступила в редакцию 02.02.2026
Reviewed / Поступила после рецензирования 03.03.2026
Accepted / Принята к публикации 20.03.2026