

УДК 519.688+004.932.2

DOI: [10.26102/2310-6018/2026.54.3.009](https://doi.org/10.26102/2310-6018/2026.54.3.009)

Вычислительный метод сегментации изображений на основе поля Дирихле и анализ асимптотической точности дискретизации пространственных регуляризаторов

Е.Ю. Щетинин✉, А.А. Шевчук

Севастопольский государственный университет, Севастополь, Российская Федерация

Резюме. Предложен вычислительный метод семантической сегментации изображений с оценкой распределительной неопределенности на основе представления предсказания в виде поля распределений Дирихле. В отличие от подходов, требующих многократных стохастических прогонов при инференсе (MC-dropout) или усреднения по ансамблю независимых моделей, метод вычисляет карты неопределенности в замкнутой форме по параметрам поля Дирихле, предсказанным за один прямой проход нейросети. Метод формулируется как минимизация составного функционала, включающего ожидаемую логарифмическую функцию потерь (expected log-loss), KL-регуляризацию для управления концентрацией распределения и пространственное сглаживание, учитывающее локальные перепады интенсивности изображения (edge-aware). Для фиксированных гладких полей установлена асимптотическая точность дискретизации используемых пространственных регуляризаторов: дискретная энергия Дирихле аппроксимирует соответствующий непрерывный интеграл с погрешностью первого порядка по шагу сетки. Дополнительно введено формальное разложение общей неопределенности на эпистемическую и подтвержденную данными компоненты, которое может использоваться в дальнейшем при анализе поведения метода и построении расширений. Вычислительные эксперименты выполнены на трех наборах медицинских изображений (ACDC, Synapse, CHAOS) с 10 независимыми инициализациями. В основном сравнении с базовой моделью, обученной по кросс-энтропии, различия статистически значимы по инициализациям на всех датасетах; на ACDC дополнительно подтверждена значимость на уровне пациентов. Метод повышает качество сегментации и улучшает калибровку вероятностных оценок при накладных расходах порядка 17%. В задаче детекции ошибок сегментации на уровне пикселей карта неопределенности достигает AUROC 0,891.

Ключевые слова: сегментация изображений, нейросетевые методы, распределение Дирихле, оценка неопределенности, калибровка, энергия Дирихле, edge-aware регуляризация, асимптотическая точность дискретизации.

Благодарности: Работа выполнена при финансовой поддержке Севастопольского государственного университета, проект 42-01-09/319/2025-1.

Для цитирования: Щетинин Е.Ю., Шевчук А.А. Вычислительный метод сегментации изображений на основе поля Дирихле и анализ асимптотической точности дискретизации пространственных регуляризаторов. *Моделирование, оптимизация и информационные технологии*. 2026;14(3). URL: <https://moitvvt.ru/ru/journal/article?id=2204> DOI: 10.26102/2310-6018/2026.54.3.009

A computational method for image segmentation based on a Dirichlet field and an analysis of the asymptotic accuracy of spatial regularizer discretization

E.Yu. Shchetinin✉, A.A. Shevchuk

Sevastopol State University, Sevastopol, the Russian Federation

Abstract. A computational method for semantic image segmentation with distributional uncertainty estimation is proposed based on representing the prediction as a Dirichlet distribution field. Unlike approaches that require multiple stochastic inference runs (MC dropout) or averaging over an ensemble of independent models, the method computes uncertainty maps in closed form based on the Dirichlet field parameters predicted in a single forward pass of the neural network. The method is formulated as the minimization of a composite functional including the expected logarithmic loss function (expected log-loss), KL regularization for controlling the distribution concentration, and spatial smoothing that takes into account local image intensity variations (edge-aware). For fixed smooth fields, the asymptotic discretization accuracy of the spatial regularizers used is established: the discrete Dirichlet energy approximates the corresponding continuous integral with a first-order error over the grid step. Additionally, a formal decomposition of the overall uncertainty into epistemic and data-supported components was introduced, which can be used in further analysis of the method's behavior and the development of extensions. Computational experiments were performed on three medical image datasets (ACDC, Synapse, CHAOS) with 10 independent initializations. In the main comparison with the baseline model trained using cross-entropy, the differences are statistically significant across initializations on all datasets; for ACDC, significance at the patient level was further confirmed. The method improves segmentation quality and improves the calibration of probability estimates with an overhead of approximately 17 %. In the task of detecting pixel-level segmentation errors, the uncertainty map achieves an AUROC of 0.891.

Keywords: image segmentation, neural network methods, Dirichlet distribution, uncertainty estimation, calibration, Dirichlet energy, edge-aware regularization, asymptotic sampling accuracy.

Acknowledgements: This work was supported by Sevastopol State University, project No. 42-01-09/319/2025-1.

For citation: Shchetinin E.Yu., Shevchuk A.A. A computational method for image segmentation based on a Dirichlet field and an analysis of the asymptotic accuracy of spatial regularizer discretization. *Modeling, Optimization and Information Technology*. 2026;14(3). (In Russ.). URL: <https://moitvvt.ru/journal/article?id=2204> DOI: 10.26102/2310-6018/2026.54.3.009

Введение

Задача семантической сегментации медицинских изображений требует не только точного предсказания классов пикселей, но и надежной оценки неопределенности результатов [1, 2]. В клинической практике карты неопределенности позволяют врачу идентифицировать области, требующие дополнительного внимания, и повышают доверие к автоматизированным системам диагностики [3]. Особую важность оценка неопределенности приобретает при работе с данными, отличающимися от обучающей выборки (out-of-distribution, OOD), где модель должна корректно сигнализировать о снижении уверенности в предсказаниях.

Существующие методы оценки неопределенности можно разделить на две основные группы. Первую группу составляют методы, основанные на стохастическом сэмплинге: MC-dropout [4], глубокие ансамбли [5], SWAG [6]. Эти методы требуют многократного выполнения прямого прохода сети (обычно $T = 10-50$ раз для MC-dropout или $M = 5-10$ моделей для ансамблей), что существенно увеличивает вычислительные затраты на этапе инференса. Например, для обработки одного изображения размером 256×256 пикселей MC-dropout с $T = 30$ проходами требует в 30 раз больше времени, чем однократный проход, что делает метод непригодным для интерактивных приложений.

Вторую группу составляют методы прямого предсказания параметров распределения: Evidential Deep Learning [7], Prior Networks [8], Posterior Networks [9]. Эти методы обучают нейросеть предсказывать параметры распределения над классами, из которых неопределенность вычисляется аналитически за один проход сети. Однако существующие реализации не включают пространственную регуляризацию, что

приводит к пространственно несогласованным картам неопределенности с характерными «шахматными» артефактами [10].

Вычислительная сложность методов первой группы составляет $O(N \cdot T)$, где N – число пикселей, T – число стохастических проходов. Для интерактивных медицинских приложений, требующих обработки в реальном времени (менее 100 мс на изображение как инженерный ориентир), такие затраты могут быть ограничением [11]. Методы второй группы обеспечивают оценку неопределенности за $O(N)$ операций, однако качество получаемых карт неопределенности часто уступает ансамблевым методам.

В настоящей работе предлагается вычислительный метод сегментации, сочетающий вероятностное представление предсказаний и вариационную регуляризацию. Нейросеть формирует в каждом пикселе параметры распределения Дирихле, а обучение задается минимизацией функционала, включающего три слагаемых: 1) ожидаемую логарифмическую функцию потерь (expected log-loss), согласующую распределительное предсказание с разметкой; 2) KL-регуляризацию, ограничивающую концентрацию распределения и тем самым подавляющую избыточную уверенность модели; 3) пространственное сглаживание, учитывающее локальные перепады интенсивности изображения (edge-aware), для повышения согласованности предсказаний и карт неопределенности без размывания границ объектов. По структуре такая постановка близка к классической регуляризации Тихонова для некорректных задач: слагаемое согласования с данными дополняется стабилизирующими регуляризаторами, контролирующими гладкость и масштаб неопределенности¹.

Основной вклад работы составили следующие положения:

1. Вычислительная схема. Предложена практическая схема минимизации составного функционала, реализуемая стандартными методами стохастической оптимизации; при фиксированной архитектуре трудоемкость одного шага обучения масштабируется линейно по числу пикселей N (с константой, зависящей от глубины/ширины сети и числа каналов). По профилированию накладные расходы относительно обучения по кросс-энтропии составляют порядка 17 %.

2. Постановка функционала. Сформулирован функционал обучения, структурно аналогичный регуляризации Тихонова: слагаемое согласования с данными (ожидаемая логарифмическая функция потерь) дополняется KL-регуляризацией, управляющей концентрацией распределения Дирихле, и пространственными регуляризаторами энергии Дирихле, учитывающими локальные перепады интенсивности изображения, – для матожидания t и логарифма концентрации $s = \log S$.

3. Теоретическое обоснование дискретизации. Доказана асимптотическая точность дискретизации используемых пространственных регуляризаторов энергии Дирихле: для фиксированных гладких полей дискретная форма аппроксимирует соответствующий непрерывный интеграл с погрешностью первого порядка по шагу сетки h .

4. Экспериментальная проверка. Проведены вычислительные эксперименты на трех публичных наборах медицинских изображений (ACDC, Synapse, CHAOS) с 10 независимыми инициализациями, включающие сравнение с базовой моделью, обученной по кросс-энтропии, с комбинацией кросс-энтропии и Dice-потерь, а также с методами многократного стохастического инференса (MC-dropout) и ансамблирования моделей.

5. Абляционный анализ. Выполнен абляционный анализ, количественно оценивающий вклад отдельных компонентов функционала (KL-регуляризации и пространственных слагаемых) в метрику качества сегментации и калибровки.

¹ Тихонов А.Н., Арсенин В.Я. *Методы решения некорректных задач*. Москва: Наука; 1979. 285 с.

6. Оценка карт неопределенности. Показано, что построенная карта неопределенности применима для ранжирования пикселей в задаче обнаружения ошибок сегментации; в выбранном протоколе оценки получено значение AUROC = 0,891.

Материалы и методы

Рассмотрим равномерную дискретизацию области Ω узлами Ω_h с шагом h . Для аппроксимации пространственных производных введем направленные разности вперед D_1^+, D_2^+ (и D_3^+ для 3D), что соответствует стандартной 4-связности в 2D и обеспечивает корректный предельный переход при $h \rightarrow 0$. Входные данные состоят из изображения $I: \Omega_h \rightarrow R^C$ (C – число каналов) и частичной разметки $y: \Omega_L \rightarrow \Delta^{K-1}$, где $\Omega_L \subseteq \Omega_h$ – размеченные пиксели, K – число классов сегментации, Δ^{K-1} – стандартный $(K-1)$ -симплекс. Задача состоит в предсказании класса каждого пикселя вместе с оценкой неопределенности предсказания. Таблица 1 содержит основные обозначения, используемые в работе.

Таблица 1 – Основные обозначения

Table 1 – Main symbols

Обозначение	Описание	Значения
Ω_h	Дискретная область (пиксели)	256×256
h	Шаг дискретной сетки	1–8 пикселей
K	Число классов сегментации	4 (ACDC), 9 (Synapse)
$\alpha(x) \in R^K_{>0}$	Параметры Дирихле	$\alpha_k \in [\varepsilon, S_{max}/K]$
$S(x) = \sum_k \alpha_k(x)$	Сумма концентрации	$S \in [1, 10^3]$
$m(x) = \alpha(x)/S(x)$	Предиктивное среднее (ожидаемые вероятности классов)	$m \in \Delta^{K-1}$
$U(x)$	Карта неопределённости	$U \geq 0$

Для каждого пикселя $x \in \Omega_h$ зададим распределение Дирихле $Dir(\pi(x) | \alpha(x))$ с параметрами концентрации $\alpha(x) \in R^K_{>0}$. Плотность распределения имеет вид:

$$p(\pi | \alpha) = (1/B(\alpha)) \prod_{k=1}^K \pi_k^{\alpha_k - 1},$$

где $B(\alpha) = \prod_k \Gamma(\alpha_k) / \Gamma(\sum_k \alpha_k)$ – мультиномиальная бета-функция.

Распределение Дирихле является сопряженным priorом для категориального распределения, что обеспечивает интерпретируемость параметров. Сумма концентрации $S(x) = \sum_k \alpha_k(x)$ определяет общую уверенность модели: при $S \rightarrow \infty$ распределение концентрируется около $m(x) = \alpha(x)/S(x)$, при малых S – близко к равномерному на симплексе. Функция $m(x)$ интерпретируется как предсказание вероятностей классов. Параметры $\alpha(x) = f_\theta(I)(x)$ предсказываются нейросетью f_θ с выходным слоем:

$$\alpha_k(x) = \text{softplus}(z_k(x)) + \varepsilon,$$

где $\varepsilon = 10^{-3}$, $z(x) = f_\theta(I)(x)$ – выход сети до активации. Функция *softplus* обеспечивает гладкость и положительность, добавка ε гарантирует строгую положительность и численную устойчивость.

В качестве карты неопределенности используем взаимную информацию в иерархии Cat–Dir, которая отражает неопределенность именно по метке Y при наличии распределения над вероятностями p . Предиктивное распределение по меткам задается математическим ожиданием:

$$m(x) = E_{p \sim Dir(\alpha(x))}[p] = \alpha(x)/S(x),$$

где $S(x) = \sum_{k=1}^K \alpha_k(x)$. Определим карту неопределенности как

$$U(x) = I(Y; p | \alpha(x)) = H(m(x)) - E_{p \sim \text{Dir}(\alpha(x))}[H(p)],$$

где $H(\cdot)$ – энтропия категориального распределения. Для $\text{Dir}(\alpha)$ второй член вычисляется аналитически:

$$E_{p \sim \text{Dir}(\alpha)}[H(p)] = \sum_{k=1}^K \left(\frac{\alpha_k}{S}\right) \cdot (\psi(S+1) - \psi(\alpha_{k+1})),$$

где $\psi(\cdot)$ – дигамма-функция. Эта величина используется как оценочный показатель (score) для выявления ошибок и построения кривых «риск-покрытие» (risk-coverage) (чем выше U , тем более неоднозначно предсказание). Дискретный функционал качества имеет вид:

$$J_h[\theta] = (1/|\Omega_L|) \sum_{x \in \Omega_L} L_{data}(\alpha(x), y(x)) + \lambda_1 R_{KL}[\alpha] + \lambda_2 R_D[m] + \lambda_3 R_D[s], \quad (1)$$

где компоненты определяются следующим образом:

– Слагаемое согласования с данными (Data term): математическое ожидание кросс-энтропии (логарифмической функции потерь expected cross-entropy (log-loss)) при истинном классе y : $L_{data} = -E_{\pi \sim \text{Dir}(\alpha)}[\log \pi_y] = \psi(S) - \psi(\alpha_y)$. Этот член штрафует расхождение между предсказанной модой $m(x)$ и истинной разметкой $y(x)$. В отличие от стандартной кросс-энтропии (cross-entropy), ожидаемая log-loss функция потерь учитывает распределение по вероятностям классов, а не только точечную оценку.

– KL-регуляризация:

$$R_{KL} = (1/N_h) \sum_{x \in \Omega_h} KL(\text{Dir}(\alpha(x)) || \text{Dir}(1))$$

ограничивает отклонение от равномерного prior $\text{Dir}(1)$, предотвращая чрезмерную уверенность (overconfidence) модели на неразмеченных или сложных пикселях. Коэффициент λ_1 контролирует баланс между точностью предсказаний и регуляризацией концентрации.

Для обеспечения пространственной гладкости (регулярности) полей m и $s = \log S$ вводятся квадратичные регуляризаторы типа энергии Дирихле. Пусть $\Omega \subset \mathbb{R}^d$ – ограниченная область, $\Omega_h = \Omega \cap (h\mathbb{Z}^d)$ – равномерная прямоугольная сетка с шагом $h > 0$, а узлы задаются как $x_i = hi$, $i \in I_h \subset \mathbb{Z}^d$. Для $k = 1, \dots, d$ обозначим через e_k единичный сдвиг на один шаг сетки вдоль оси k и введем нормированные прямые конечные разности $D_k^+ u_i := \frac{u_{i+e_k} - u_i}{h}$, $i, i + e_k \in I_h$. Суммирование далее ведется по тем i , для которых разности определены (например, по «внутренним» узлам относительно направления k).

Энергия Дирихле для векторного поля m . Пусть $m_i \in \mathbb{R}^K$ – узловые значения поля параметров по классам. Тогда регуляризатор задается как

$$R_{D,h}[m] := \frac{h^d}{2} \sum_{i \in I_h} \sum_{k=1}^d \omega_{i,k} \|D_k^+ m_i\|_2^2, \quad (2)$$

где $\|\cdot\|_2$ – евклидова норма в \mathbb{R}^K , а $\omega_{i,k}$ – вес на ребре $(i, i + e_k)$.

Дискретная энергия Дирихле (edge-aware) для скалярного поля s . Пусть $s_i \in \mathbb{R}$ – узловые значения скалярного поля $s = \log S$. Тогда

$$R_{D,h}[s] := \frac{h^d}{2} \sum_{i \in I_h} \sum_{k=1}^d \omega_{i,k} |D_k^+ s_i|^2. \quad (3)$$

В двумерном случае $d = 2$ сумма берётся по $k \in \{1, 2\}$, а множитель h^d равен h^2 при использовании нормированных разностей D_k^+ .

Edge-aware веса. Термин edge-aware означает, что вес $\omega_{i,k}$ уменьшается на ребрах, вдоль которых наблюдается резкий локальный перепад интенсивности изображения I , чтобы регуляризация не «сшивала» значения m и s через предполагаемые границы объектов. Практически веса можно задавать, например, как

$$\omega_{i,k} = \max\{\omega_{min}, \exp(-\beta|D_k^+ I_i|^2)\}, \omega_{min} > 0, \beta > 0,$$

где β задает чувствительность к границам, а отсечка ω_{min} обеспечивает невырожденность (веса ограничены снизу и не стремятся к нулю).

Замечание (связь с Теоремой 1). Веса $\omega_{i,k}$ рассматриваются как дискретизация некоторой гладкой функции $\omega(x)$ на центрах ребер $x_{i+e_k/2}$ и предполагаются равномерно невырожденными; конкретная edge-aware формула – это практический выбор весов на фиксированном разрешении с гарантией $\omega_{i,k} \geq \omega_{min}$.

Для обоснования дискретизации рассматриваются фиксированные гладкие поля $m: \bar{\Omega} \rightarrow \mathbb{R}^K$ и $s: \bar{\Omega} \rightarrow \mathbb{R}$ на ограниченной прямоугольной области $\Omega \subset \mathbb{R}^d$.

Пусть задана равномерная прямоугольная сетка с шагом $h > 0$. Обозначим через $\mathcal{J}_h \subset \mathbb{Z}^d$ множество мультииндексов узлов, попадающих в Ω , и положим $x_i := hi$, $\Omega_h := \{x_i: i \in \mathcal{J}_h\}$. Узловые выборки полей обозначены как $m_i := m(x_i)$, $s_i := s(x_i)$. Соседние узлы связываются ориентированными ребрами вида $(i, i + e_k)$ – канонический вектор в \mathbb{Z}^d (сдвиг на один шаг по оси k), причем рассматриваются только те ребра, для которых $i \in \mathcal{J}_h$ и $i + e_k \in \mathcal{J}_h$. Для каждого такого ребра задан реберный вес $\omega_{i,k}$. Цель данного раздела – показать, что дискретные пространственные регуляризаторы, используемые в функционале (1), являются асимптотически точными аппроксимациями соответствующих непрерывных интегралов при $h \rightarrow 0$ с контролируемой скоростью по h . В частности, данное утверждение применяется к регуляризаторам для векторного поля m и скалярного поля s , заданным формулами (2)–(3).

Далее проведем исследование асимптотической точности дискретизации энергии Дирихле. Введем нормированные конечные разности для узлового поля $u: \Omega_h \rightarrow \mathbb{R}^r$, $D_k^h u_i := \frac{u_{i+e_k} - u_i}{h}$, $k = 1, \dots, d$, определенные для всех $i = (i_1, \dots, i_d) \in \mathcal{J}_h$, таких что $i + e_k \in \mathcal{J}_h$. Далее рассмотрим дискретную энергию Дирихле с реберными весами $\omega_{i,k}$ в следующем виде:

$$R_{D,h}[u_h] := h^d \sum_{i \in \Omega_h} \sum_{k=1}^d \omega_{i,k} \|D_k^h u_i\|_2^2, \quad (4)$$

где $\|\cdot\|_2$ – евклидова норма \mathbb{R}^r (в скалярном случае $r = 1$ совпадает с модулем). Формулы (2) и (3) являются частными случаями (4) при $u = m \in \mathbb{R}^K$ (векторное поле размерности $r = K$) и $u = s \in \mathbb{R}$ (скалярное поле $r = 1$) соответственно.

Гипотеза 1 (невырожденность и согласованность весов). Существует функция $\omega \in C^1(\bar{\Omega})$, не зависящая от h , и константы $c_0, c_1, C_\omega > 0$, такие, что для каждого ребра $(i, i + e_k)$ с центром $x_{i+e_k/2} := h(x_i + \frac{1}{2}e_k)$ при $h \rightarrow 0$ выполнено условие:

$$|\omega_{i,k} - \omega(x_{i+e_k/2})| \leq C_\omega h, c_0 \leq \omega_{i,k} \leq c_1 < \infty.$$

Замечание. В вычислительной части веса $\omega_{i,k}$ строятся по изображению I (edge-aware) и затем применяется нижняя отсечка $\omega_{i,k} \leftarrow \max\{\omega_{i,k}, \omega_{min}\}$ с ω_{min} , что обеспечивает невырожденность $c_0 \leq \omega_{i,k}$. Теорема 1 ниже относится к аппроксимационным свойствам дискретного регуляризатора (4) при весах, удовлетворяющих Гипотезе 1.

Теорема 1 (асимптотическая точность дискретизации энергии Дирихле). Пусть $u \in C^2(\bar{\Omega}; \mathbb{R}^r)$ (где $r = K$ для $u = m$ и $r = 1$ для $u = s$) и выполнена Гипотеза 1. Если $u_i = u(x_i)$ – узловая выборка на равномерной сетке Ω_h с шагом h , то справедливо:

$$R_{D,h}[u] = \int_{\Omega} \omega(x) \|\nabla u(x)\|_F^2 dx + O(h), \quad h \rightarrow 0, \quad (5)$$

где $\|\nabla u(x)\|_F^2 = \sum_{k=1}^d \|\partial_k u(x)\|_2^2$ (в скалярном случае $r = 1$ это $|\nabla u(x)|^2$).

Доказательство. Зафиксируем $k \in \{1, \dots, d\}$. По формуле Тейлора для $u \in C^2$ имеем равномерно по всем допустимым узлам i :

$$u_{i+e_k} = u_i + h\partial_k u(x_i) + O(h^2),$$

откуда $D_k^h u_i = \frac{u_{i+e_k} - u_i}{h} = \partial_k u(x_i) + O(h)$. Кроме того, из гладкости $\partial_k u$ следует:

$$\partial_k u(x_i) = \partial_k u(x_{i+e_k/2}) + O(h),$$

а и из Гипотезы 1 получаем: $\omega_{i,k} = \omega(x_{i+e_k/2}) + O(h)$, $0 \leq \omega_{i,k} \leq c_1$.

Следовательно, $\omega_{i,k} \|D_k^h u_i\|_2^2 = \omega_{i,k}(x_{i+e_k/2}) \|\partial_k u(x_{i+e_k/2})\|_2^2 + O(h)$, равномерно по i . Подставляя в (4) и суммируя по k , получаем риманову сумму для функции:

$$g(x) := \omega(x) \|\nabla u(x)\|_F^2.$$

Так как $\omega \in C^1(\bar{\Omega})$ и $u \in C^2(\bar{\Omega})$, то $g \in C^1(\bar{\Omega})$, и для соответствующей римановой суммы справедлива стандартная оценка:

$$|h^d \sum g(\cdot) - \int_{\Omega} g| \leq Ch,$$

где C зависит только от $\|g\|_{C^1(\bar{\Omega})}$ и способа суммирования по внутренним ребрам. Отсюда следует (5).

Следствие (первый порядок и экстраполяция Ричардсона). Если $R_{D,h}[u_h] = R[u] + Ch + o(h)$, то $ratio = c_h/c_{h/2} \rightarrow 2$, при $h \rightarrow 0$, т. е. наблюдаемое $ratio \approx 2$ соответствует порядку $O(h)$. В этом случае возможна апостериорная экстраполяция Ричардсона: $R^* \approx 2R_{D,h/2}[u_{h/2}] - R_{D,h}[u_h]$.

Ниже приведена эмпирическая проверка согласованности (асимптотической точности) дискретизации энергии Дирихле на синтетическом гладком поле u . Для последовательности равномерных сеток Ω_h с шагом h (от $h = 8$ до $h = 0,5$) вычислялись дискретные значения $R_{D,h}[u_h]$ и сравнивались с эталонным значением $R[u]$, полученным аналитически (либо численно на существенно более тонкой сетке). В соответствии с Теоремой 1 наблюдается первый порядок точности по h : $e(h) := |R_{D,h}[u_h] - R[u]|$ убывает как $e(h) = O(h)$, что эквивалентно соотношению $h \rightarrow 0$. Указанное поведение отражено в Таблице 2. При таком режиме сходимости, при необходимости, может быть использована апостериорная экстраполяция Ричардсона $R[u] \approx 2R_{D,h/2}[u] - R_{D,h}[u]$ для подавления ведущей ошибки порядка $O(h)$.

Таблица 2 – Эмпирическая проверка асимптотической точности дискретизации энергии Дирихле

Table 2 – Empirical verification of the asymptotic accuracy of Dirichlet energy discretization

h (пикс.)	$e(h)$	$Ratio(h)$
8	–	–
4	$4,2 \cdot 10^{-2}$	–
2	$1,9 \cdot 10^{-2}$	2,21
1	$8,5 \cdot 10^{-3}$	2,24
0,5	$3,8 \cdot 10^{-3}$	2,24

Наблюдаемое $Ratio \approx 2$ согласуется с режимом $e(h) = O(h)$, предсказанным Теоремой 1. Выбор $h = 1$ (базовое разрешение) в основных экспериментах представляет практический компромисс между дискретизационной погрешностью регуляризатора $R_{D,h}[u]$ и вычислительными затратами. Дальнейшее уменьшение h приводит к умеренному снижению $e(h)$ при заметном росте потребления памяти и времени вычислений.

Минимизация функционала $J_h(\theta)$ выполняется стохастическим градиентным методом с адаптивной оценкой моментов (Adam) [12]. На каждой итерации t выполняются следующие шаги.

1. Формирование мини-пакета (mini-batch). Из обучающей выборки случайно выбирается мини-пакет B изображений размера $|B| = 8$.

2. Прямой проход сети. Для каждого изображения $I(b), b \in B$, вычисляются параметры распределения Дирихле (вектор $\alpha^{(b)}(x)$ по классам в каждом пикселе), то есть выходы Dirichlet-head нейросети.

3. Вычисление функционала. Вычисляются все слагаемые функционала: слагаемое согласования с данными L_{data} (ожидаемая логарифмическая функция потерь), которое выражается через специальные функции (в частности, дигамма-функцию).

4. KL-регуляризация R_{KL} как дивергенция Кульбака–Лейблера между предсказанным распределением Дирихле и опорным (равномерным) распределением; пространственные регуляризаторы $R_{D,h}[m]$ и $R_{D,h}[s]$, вычисляемые суммированием по ребрам графа сетки (взвешенная дискретная энергия Дирихле).

5. Вычисление градиента. Градиент $\nabla_{\theta} J_h(\theta)$ вычисляется методом обратного распространения ошибки. Все компоненты функционала дифференцируемы, что обеспечивает корректность градиентного шага.

6. Обновление параметров. Параметры обновляются по правилу Adam с шагом $\eta = 10^{-4}$ и стандартными коэффициентами $\beta_1 = 0,9$; $\beta_2 = 0,999$.

При фиксированной архитектуре (фиксированные глубина/ширина сети) трудоемкость одного шага обучения (прямой + обратный проход) линейна по числу пикселей N : $T_{\text{net}} = O(pN)$, где p – число параметров сети, а скрытая константа зависит от числа слоёв, числа каналов и размеров сверточных ядер. Дополнительные члены функционала сохраняют линейную зависимость по N . Пусть K – число классов, а $|\mathcal{E}|$ – число ребер графа сетки (для 2D $|\mathcal{E}| \approx 2N$, для 3D $|\mathcal{E}| \approx 3N$). Вычисление специальных функций (softplus, digamma) и KL-регуляризации выполняется поэлементно и требует $O(KN)$ операций. Вычисление весов ω_{ij} на ребрах требует $O(|\mathcal{E}|)$. Регуляризатор для моды m (векторное поле размерности K) требует $O(K|\mathcal{E}|)$, а для s (скалярное поле) $-O(|\mathcal{E}|)$.

Таблица 3 – Дополнительные вычислительные затраты компонентов функционала
 Table 3 – Additional computational costs of functional components

Компонент	Сложность	Накладные расходы, %
softplus + digamma	$O(K \cdot N)$	0,3
KL-дивергенция R_{KL}	$O(K \cdot N)$	0,4
Edge-aware веса ω_{ij}	$O(E)$	1,2
энергия Дирихле $R_{D,h}[m] + R_{D,h}[s]$	$O((K + 1) \cdot E)$	0,6
Итого (доп. члены, forward)	$O(K \cdot N) + O((K + 1) \cdot E)$	~2,5

Суммарные накладные расходы дополнительных слагаемых составляют около 2,5 % от времени прямого прохода. Полные накладные расходы на шаг обучения (прямой + обратный проход), измеренные профилированием, составляют около 17 %.

Численная реализация предложенного метода включает приемы, направленные на снижение потребления памяти и повышение численной устойчивости:

1. Пересчет промежуточных активаций (gradient checkpointing) [13]. Промежуточные активации энкодера частично не сохраняются и пересчитываются при обратном проходе, что уменьшает потребление GPU-памяти и позволяет обрабатывать изображения большего размера.

2. Смешанная точность (AMP) [14]. На этапе обучения используется автоматический выбор смешанной точности (autocast) и масштабирование градиентов (GradScaler), что ускоряет вычисления на современных GPU (например, A100/V100) без заметной деградации целевых метрик.

3. Ограничение концентрации S для устойчивого вычисления специальных функций. Для предотвращения численных проблем при больших аргументах вводится ограничение сверху $S \leftarrow \min(S, S_{max})$ (реализация: `torch.clamp`), при этом положительность S обеспечивается параметризацией (например, через *softplus*). Доля активаций, попадающих под ограничение сверху, составляет менее 0,1 %; по динамике функции потерь и целевых метрик заметного влияния ограничения не наблюдалось. Параметры по умолчанию (включая S_{max} и коэффициенты регуляризации) выбирались по результатам анализа чувствительности/абляционного исследования.

Численные эксперименты проведены на трех общедоступных наборах медицинских изображений, широко используемых для сопоставления методов семантической сегментации.

ACDC [15] – сегментация структур сердца на МРТ (левый желудочек LV, правый желудочек RV, миокард Myo). Набор содержит $n = 100$ пациентов и $K = 4$ класса (включая фон). Разбиение выполнено по пациентам в пропорции 70/10/20 (train/val/test). Данные получены на сканерах 1,5T и 3T с различающимися протоколами, что обеспечивает существенную вариабельность изображений.

Synapse² – мультиорганная сегментация брюшной полости на КТ. Набор содержит $n = 30$ случаев и $K = 9$ классов (селезенка, почки, печень, желудок, аорта, нижняя полая вена, поджелудочная железа, надпочечники и др.). Сложность обусловлена большим числом классов и вариабельностью анатомии.

CHAOS [16] – сегментация печени на МРТ. Набор содержит $n = 40$ случаев и $K = 4$ класса. Используются последовательности T1-DUAL и T2-SPiR, что дополнительно увеличивает вариабельность контраста.

Для оценки вариабельности результатов и статистической значимости различий между методами использовались 10 независимых запусков (seeds, случайных инициализаций).

Метрики качества сегментации:

1. Коэффициент Dice – мера перекрытия предсказанной маски A и разметки B:
$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$
 В экспериментах Dice усредняется по классам (macro-average) и по изображениям тестовой выборки.

2. Ошибка калибровки ECE (Expected Calibration Error) [17] – мера согласованности уверенности модели и фактической точности. Пиксели группируются в $B = 15$ равноширинных интервалов (бинов) по предсказанной уверенности; для каждого

² Landman B., Xu Zh., Igelsias J., et al. *MICCAI Multi-Atlas Labeling Beyond the Cranial Vault (Cervix)*. Hugging Face. URL: https://huggingface.co/datasets/loupk/CADS-dataset/blob/main/0006_bcv_cervix/README_0006_bcv_cervix.md (дата обращения: 19.01.2026).

бина вычисляется разность между средней уверенностью и эмпирической точностью, после чего берется взвешенная сумма по бинам.

3. AUROC для выявления ошибок сегментации – вероятность корректного ранжирования случайной пары пикселей («ошибочный»/«корректный») по значению неопределенности $U(x)$. Положительный класс соответствует пикселям с ошибкой сегментации (предсказание \neq разметка).

4. AUPRC (площадь под кривой precision–recall) – более информативная метрика при дисбалансе классов. В рассматриваемых экспериментах доля ошибочных пикселей составляет порядка 8 %, то есть положительный класс существенно менее представлен, чем отрицательный.

Все метрики качества (Dice, ECE) вычислялись на тестовой выборке и агрегировались на уровне пациентов: сначала метрика усреднялась по срезам/пикселям внутри одного исследования, затем выполнялось усреднение по пациентам (macro-average по пациентам). При расчете AUROC/AUPRC для выявления ошибок «ошибочным» считался пиксель, для которого предсказанный класс не совпадает с разметкой; пиксели с неопределенной разметкой (void, если присутствуют) исключались из расчета. Фоновый класс учитывался наравне с остальными, если не указано иное, что обеспечивает воспроизводимость при различной доле фона в изображениях.

В работе проведен сравнительный анализ со следующими методами, представляющими разные подходы к оценке неопределенности. SE: обучение по стандартной кросс-энтропии без явного механизма неопределенности; оценка неопределенности – энтропия softmax-вероятностей. SE+Dice: сумма кросс-энтропии и функции потерь Dice с равными весами; распространенная комбинация в медицинской сегментации. MC-dropout [4]: $T = 30$ стохастических прямых проходов с включенным dropout (вероятность $p = 0,5$) на этапе инференса; неопределенность оценивается по разбросу предсказаний между проходами. Глубокие ансамбли [5]: $M = 5$ моделей одной и той же архитектуры (U-Net/ResNet-34) с одинаковыми гиперпараметрами, различающихся только инициализацией весов; неопределенность оценивается по разбросу предсказаний между моделями.

Архитектура. Используется U-Net [18] с энкодером ResNet-34 [19], предварительно обученным на ImageNet. Декодер состоит из четырех блоков с пропусками (skip-connections) и билинейным повышением разрешения. Число параметров сети составляет $p \approx 24,4$ млн.

Протокол обучения. Оптимизатор Adam, шаг обучения $\eta = 10^{-4}$, 200 эпох, размер мини-пакета 8. Ранняя остановка по Dice на валидации (patience 20 эпох). Аугментации: случайные повороты ($\pm 15^\circ$), масштабирование (0,9–1,1), горизонтальное отражение. Размер входного изображения: 256×256 .

Замечание (корректность сравнения с ансамблями). Для сопоставимости ансамбль формируется из $M = 5$ моделей той же архитектуры и с теми же гиперпараметрами, что и предлагаемый метод; различие состоит только в случайной инициализации. Это обеспечивает корректное сравнение по архитектуре, при этом суммарные затраты на обучение ансамбля возрастают примерно в M раз.

Результаты по качеству сегментации приведены в Таблице 4. Во всех трех наборах данных предлагаемый метод показывает улучшение относительно базовой модели SE при сопоставимой дисперсии результатов по 10 независимым инициализациям.

Таблица 4 – Качество сегментации
Table 4 – Segmentation quality

Метод	ACDC	Synapse	CHAOS
CE	0,891 ± 0,008	0,762 ± 0,012	0,883 ± 0,009
CE+Dice	0,904 ± 0,006	0,779 ± 0,010	0,895 ± 0,007
MC-dropout ($T = 30$)	0,897 ± 0,007	0,771 ± 0,010	0,889 ± 0,008
Ансамбли ($M = 5$)	0,908 ± 0,005	0,785 ± 0,008	0,901 ± 0,006
Предлагаемый метод	0,912 ± 0,005	0,789 ± 0,007	0,906 ± 0,005

Статистическая значимость. Для каждого датасета отдельно выполнен двусторонний парный критерий Уилкоксона по 10 значениям Dice, полученным при 10 различных инициализациях на фиксированном тестовом наборе (сравнение предлагаемый метод vs «CE (базовый)» отдельно для каждого датасета). Для учета множественных проверок по трем датасетам применялась поправка Бонферрони ($\min(3p, 1)$); получены скорректированные значения: ACDC $p \approx 0,002$ ($p_{adj} = 0,006$), Synapse $p \approx 0,002$ ($p_{adj} = 0,006$), CHAOS $p \approx 0,002$ ($p_{adj} = 0,006$). Для набора ACDC дополнительно выполнена проверка на уровне пациентов: для каждого пациента тестового набора ($n = 20$) вычислялся Dice с агрегацией по всем срезам/объему, после чего применялся двусторонний парный критерий Уилкоксона. Получено значение $p < 0,001$, что подтверждает устойчивость улучшения на уровне пациентов и исключает объяснение эффекта только вариабельностью по инициализациям. Для Synapse ($n = 6$) и CHAOS ($n = 8$) пациент-ориентированная проверка рассматривается как качественная из-за малого размера тестового набора, при этом значимость на уровне инициализаций сохраняется.

В среднем предлагаемый метод дает качество, сопоставимое с глубокими ансамблями при меньших вычислительных затратах: обучение одной модели занимает 3,5 ч против 15,0 ч для ансамбля из $M = 5$ моделей (Таблица 3, Таблица 7, GPU A100). Наблюдаемое улучшение качества согласуется с введением в функционал двух дополнительных механизмов: KL-регуляризации, контролирующей концентрацию параметров распределения Дирихле и стабилизирующей обучение, и пространственной регуляризации с реберными весами ω_{ij} , ослабляющей сглаживание в областях с большим локальным перепадом интенсивности.

Далее в работе были проведены исследования по оцениванию качества калибровки. Результаты калибровки представлены в Таблице 5. Во всех трех датасетах предлагаемый метод снижает ошибку калибровки ECE (6) по сравнению с CE и CE+Dice, что указывает на более согласованное соответствие предсказанной уверенности и эмпирической точности.

Определение ECE. Метрика ECE вычислялась в многоклассовой постановке по схеме one-vs-rest. Для каждого класса $k \in \{1, \dots, K\}$ рассматривались предсказанные вероятности $p_k(x)$ и бинарные индикаторы истинной метки $1\{y(x) = k\}$. Значения $p_k(x)$ разбивались на $B = 15$ равноширинных бинов $B\{\mathcal{J}_b\}_{b=1}^B$ на отрезке $[0, 1]$. Для каждого бина вводились $\text{conf}_{k,b} = \frac{1}{|\mathcal{S}_{k,b}|} \sum_{x \in \mathcal{S}_{k,b}} p_k(x)$, $\text{acc}_{k,b} = \frac{1}{|\mathcal{S}_{k,b}|} \sum_{x \in \mathcal{S}_{k,b}} 1\{y(x) = k\}$, где $\mathcal{S}_{k,b} = \{x: p_k(x) \in \mathcal{J}_b\}$. Тогда поклассовая ошибка калибровки определяется как

$$\text{ECE}_k = \sum_{b=1}^B \frac{|\mathcal{S}_{k,b}|}{|\Omega|} |\text{acc}_{k,b} - \text{conf}_{k,b}|, \quad (6)$$

а итоговая метрика определяется усреднением по классам: $\text{ECE} = \frac{1}{K} \sum_{k=1}^K \text{ECE}_k$. В Таблице 5 приведены значения ECE (6) в виде $\text{mean} \pm \text{std}$ по 10 инициализациям.

Таблица 5 – Калибровка
Table 5 – Calibration

Метод	ACDC	Synapse	CHAOS
CE	0,078 ± 0,010	0,081 ± 0,012	0,092 ± 0,014
CE+Dice	0,065 ± 0,009	0,068 ± 0,010	0,079 ± 0,012
MC-dropout	0,032 ± 0,006	0,035 ± 0,007	0,048 ± 0,009
Ансамбли ($M = 5$)	0,024 ± 0,004	0,027 ± 0,005	0,039 ± 0,007
Предлагаемый	0,021 ± 0,004	0,024 ± 0,005	0,036 ± 0,006

На наборе ACDC метрика ECE уменьшается с 0,078 до 0,021. Устойчивость вывода к способу биннинга проверялась заменой равношириного биннинга на равнонаполненный (equal-mass) по каждому классу: на ACDC получено $ECE_{\text{equal-mass}} = 0,23$ против $ECE_{\text{equal-width}} = 0,021$; качественные выводы при этом не меняются. Улучшение калибровки означает, что для каждого класса k предсказанные вероятности $p_k(x)$ в среднем согласуются с эмпирической частотой события $\{y(x) = k\}$ в соответствующих интервалах уверенности, что важно при интерпретации вероятностных оценок в прикладных задачах медицинской сегментации. Снижение ECE согласуется с тем, что обучение ведется по ожидаемой логарифмической функции потерь для распределения Дирихле, то есть учитывается не точечная вероятность класса, а распределение по вероятностям, параметризованное $\alpha(x)$. Дополнительная KL-регуляризация по концентрации $S(x)$ ограничивает вырожденные режимы избыточной уверенности и тем самым поддерживает более согласованные вероятностные оценки в смысле калибровки.

Качество карт неопределенности оценивалось в задаче выявления ошибок сегментации на уровне пикселей. Каждому пикселю x сопоставлялась скалярная величина $U(x)$, используемая как оценочный показатель для ранжирования: для предлагаемого метода в качестве $U(x)$ применялась взаимная информация в иерархии Cat-Dir (sampling-free), для базового CE – энтропия softmax-распределения. Для каждого пикселя определялась бинарная метка ошибки:

$$\varepsilon(x) = 1\{\hat{y}(x) \neq y(x)\}, \quad \hat{y}(x) = \operatorname{argmax}_k \hat{p}_k(x),$$

где $y(x)$ – истинная метка, $\hat{p}_k(x)$ – предсказанная вероятность класса k (для предлагаемого метода средняя вероятность, индуцируемая $\alpha(x)$, для CE – softmax-вероятность). Пиксели с неопределенной разметкой (void, если присутствуют) исключались; фоновый класс учитывался. По ранжированию пикселей по $U(x)$ вычислялись метрики AUROC и AUPRC; в Таблице 6 приведены значения $\text{mean} \pm \text{std}$ по 10 инициализациям ($AUROC/AUPRC \pm \text{std}$; среднее по 10 запускам; время инференса на одно изображение, мс).

Таблица 6 – Качество карт неопределенности при выявлении ошибок сегментации
Table 6 – Quality of uncertainty maps when detecting segmentation errors

Метод	AUROC	AUPRC	Инференс, мс
CE (энтропия softmax)	0,812 ± 0,015	0,284 ± 0,021	22
MC-dropout ($T = 30$)	0,847 ± 0,012	0,331 ± 0,018	660
Ансамбли ($M = 5$)	0,875 ± 0,009	0,379 ± 0,015	110
Предлагаемый	0,891 ± 0,008	0,402 ± 0,014	26

Значение AUROC интерпретируется как вероятность правильного ранжирования случайной пары пикселей «ошибочный/корректный»: $AUROC = \mathbb{P}(U(x_{\text{err}}) > U(x_{\text{ok}}))$, $\varepsilon(x_{\text{err}}) = 1$, $\varepsilon(x_{\text{ok}}) = 0$. Как видно из Таблицы 6, предлагаемый метод дает более

высокие значения AUROC/AUPRC по сравнению с энтропией softmax и близкие по порядку значения к ансамблям, сохраняя при этом однопроходный режим инференса: 26 мс против 110 мс для $M = 5$ моделей (Таблица 7).

Примечание. Доля ошибочных пикселей в задаче $\varepsilon(x) \in \{0,1\}$ составляет $\pi_{\text{err}} = \frac{1}{|\Omega|} \sum_x \varepsilon(x)$ (в наших экспериментах порядка нескольких процентов), поэтому базовый уровень AUPRC для случайного ранжирования равен π_{err} . В этих условиях AUPRC является более чувствительной метрикой качества ранжирования редкого положительного класса, чем AUROC. В иллюстративных примерах значения $U(x)$ обычно повышены в областях ошибок и в зонах неоднозначности на границах; пространственная регуляризация при этом подавляет высокочастотные артефакты, характерные для несглаженных оценок.

В Таблице 7 приведены суммарные вычислительные затраты сравниваемых методов на GPU NVIDIA A100 при batch size = 8 и фиксированном протоколе обучения/инференса. Время обучения указано как суммарное время до завершения обучения одной модели (для ансамблей – суммарно по M моделям), время инференса – среднее время обработки одного изображения (для MC-dropout и ансамблей – с учетом всех прогонов/моделей), потребление памяти – пиковое значение GPU-памяти в ходе обучения.

Таблица 7 – Вычислительные затраты
 Table 7 – Computational costs

Метод	Обучение, ч	Инференс, мс	Память, GB	Overhead
CE (базовый)	3,0	22	7,6	–
CE+Dice	3,1	22	7,6	+3 %
MC-dropout ($T = 30$)	3,0	660	7,6	+2900 % (инференс)
Ансамбли ($M = 5$)	15,0	110	38,0	+400 %
Предлагаемый	3,5	26	8,2	+17 %

Предлагаемый метод добавляет умеренные накладные расходы: около +17 % к времени обучения и +18 % к времени инференса относительно CE baseline (26 мс против 22 мс на изображение при указанной конфигурации). Для сопоставления с альтернативами отметим: MC-dropout требует $T = 30$ прогонов на инференсе, что приводит к увеличению времени примерно в T раз (660 мс против 22 мс) при неизменной памяти одной модели; ансамбли требуют обучения $M = 5$ независимых моделей, что увеличивает суммарное время обучения примерно в M раз (15,0 ч против 3,0 ч) и, при одновременном размещении моделей на GPU, приводит к росту используемой памяти (38,0 GB против 8,2 GB); предлагаемый метод обеспечивает улучшенные показатели качества карт неопределенности при существенно меньших вычислительных затратах по сравнению с многопроходными подходами, сохраняя малое время инференса.

Для количественной оценки вклада отдельных слагаемых функционала выполнен абляционный анализ на датасете ACDC. Во всех конфигурациях фиксировались архитектура сети, предобработка, разбиение на обучающую/валидационную/тестовую выборки и протокол обучения; различались только включаемые члены регуляризации и соответствующие коэффициенты. В Таблице 8 приведены значения Dice и ECE ($mean \pm std$ по 10 запускам), а также NLL (усреднение по 10 запускам). Здесь под NLL понимается эмпирическая отрицательная лог-вероятность истинной метки под предиктивным категориальным распределением $\hat{p}(x) = \mathbb{E}[p(x) | \alpha(x)] = \alpha(x)/S(x)$:

$$NLL = -\frac{1}{|T|} \sum_{x \in T} \sum_{y(x)} \log. \quad (7)$$

В то же время слагаемое L_{data} в функционале (1) является expected log-loss под распределением Дирихле и выражается через дигамма-функции; эти величины совпадают лишь в частных предельных режимах.

Таблица 8 – Абляционный анализ на ACDC
Table 8 – Ablation analysis on ACDC

Конфигурация	Dice	ECE	NLL
CE (базовый)	0,891 ± 0,008	0,078 ± 0,010	0,342
+ R_{KL}	0,897 ± 0,007	0,052 ± 0,008	0,305
+ R_{KL} + $R_{D,h}[m]$	0,899 ± 0,007	0,048 ± 0,007	0,295
+ R_{KL} + $R_{D,h}[m]$ + $R_{D,h}[s]$	0,905 ± 0,006	0,035 ± 0,006	0,268
Полная модель	0,912 ± 0,005	0,021 ± 0,004	0,245

Анализ вклада компонентов.

1. Добавление R_{KL} дает наибольший вклад в улучшение калибровки: ECE снижается с 0,078 до 0,052, одновременно наблюдается рост *Dice* (0,891 → 0,897) и уменьшение NLL (0,342 → 0,305). Это согласуется с ролью R_{KL} как регуляризатора, ограничивающего избыточную уверенность и стабилизирующего вероятностные оценки.

2. Добавление пространственной регуляризации моды $R_{D,h}[m]$ приводит к дополнительному, хотя и умеренному улучшению *Dice* (0,897 → 0,899) и ECE (0,052 → 0,048), что можно интерпретировать как повышение пространственной согласованности предсказаний за счет сглаживания поля $m(x)$ на соседних узлах сетки.

3. Регуляризация лог-концентрации $R_{D,h}[s]$ дает заметный дополнительный эффект: *Dice* увеличивается до 0,905, ECE снижается до 0,035, NLL снижается до 0,268. Это указывает на практическую ценность пространственно согласованного поля $s(x) = \log S(x)$ как компонента вероятностной части модели.

4. Полная модель использует выбранные по результатам абляционного анализа значения гиперпараметров $\lambda_1 = 0,1$, $\lambda_2 = \lambda_3 = 0,01$, $\beta = 5,0$, что обеспечивает дополнительное улучшение по всем метрикам (*Dice* = 0,912, ECE = 0,021, NLL = 0,245) по сравнению с конфигурацией без итоговой настройки.

Примечание о NLL и expected log-loss. В работе термин expected log-loss используется для обозначения обучающего слагаемого согласования с данными L_{data} , вычисляемого как математическое ожидание отрицательного логарифма правдоподобия по распределению над вероятностями классов, индуцируемому параметрами Дирихле $\alpha(x)$. В то же время NLL в Таблице 8 приводится как оценочная метрика на тестовой выборке: среднее по пикселям/изображениям значение отрицательного логарифма предсказанной вероятности истинного класса (то есть стандартный отрицательный логарифм правдоподобия для наблюдаемых меток). Таким образом, expected log-loss относится к оптимизируемому функционалу, тогда как NLL используется для внешней количественной оценки качества вероятностных предсказаний.

Для оценки чувствительности метода к гиперпараметрам выполнен анализ в окрестности значений по умолчанию на датасетах ACDC/Synapse/CHAOS. Во всех сериях экспериментов фиксировались архитектура, протокол обучения и разбиения; варьировались только коэффициенты регуляризации (например, $\lambda_1, \lambda_2, \lambda_3$) и параметр построения edge-aware весов (например, β). В исследованных диапазонах существенной деградации качества не наблюдалось: изменения *Dice* оставались в пределах

естественной вариабельности по 10 запускам. Наблюдались следующие тенденции (по Dice, усреднение по 10 seeds): $\lambda_1 \in [0,01; 1]$: разброс значений метрики Dice не превышал 0,010, максимум достигался вблизи $\lambda_1 \approx 0,1$; $\lambda_2, \lambda_3 \in [0,001, 0,1]$: разброс Dice не превышал 0,012, устойчивый диапазон – около 0,01; $\beta \in [1, 50]$: разброс Dice не превышал 0,015, наилучшие значения наблюдались при $\beta \approx 5$. В наших экспериментах метод демонстрирует робастность в диапазонах $\lambda_1 \in [0,05; 0,2]$ и $\lambda_2 \in [0,005; 0,05]$ (при $\lambda_3 = \lambda_2$), что позволяет рекомендовать значения по умолчанию: $\lambda_1 = 0,1$, $\lambda_2 = \lambda_3 = 0,01$, $\beta = 5,0$. Верхняя отсечка концентрации $S3_{max}$ обеспечивает численную устойчивость вычислений специальных функций в наших экспериментах на ACDC; перенос этого значения на другие датасеты следует рассматривать как эвристику и при необходимости подтверждать отдельной проверкой.

Обсуждение

Полученные результаты показывают, что предложенная постановка на основе параметризации распределения Дирихле и регуляризации по полям m и $s = \log S$ в рассматриваемых экспериментах приводит к улучшению качества сегментации, калибровки вероятностных оценок и информативности карт неопределенности при умеренных вычислительных накладных расходах. Ниже обсуждаются основные наблюдения, их интерпретация в рамках выбранной модели, а также ограничения проведенной валидации.

Качество сегментации. Согласно Таблице 4, предлагаемый метод демонстрирует более высокие значения Dice по сравнению с базовой моделью SE на всех трех датасетах, а также значения Dice, близкие по порядку к результатам ансамблей при учете стандартного отклонения по инициализациям. Статистическая значимость улучшений оценивалась двусторонним парным критерием Уилкоксона по 10 запускам отдельно для каждого датасета с поправкой на множественные сравнения по трем датасетам. Дополнительная проверка на уровне пациентов для ACDC ($n = 20$) подтверждает сохранение эффекта при агрегации метрики по пациентам, что снижает вероятность объяснения улучшения исключительно вариативностью по инициализациям. Наблюдаемое улучшение согласуется с абляционным анализом (Таблица 8): KL-регуляризация R_{KL} стабилизирует вероятностные оценки, а пространственные члены $R_{D,h}[m]$ и $R_{D,h}[s]$ усиливают пространственную согласованность соответствующих полей. В совокупности это снижает локальные несогласованности предсказаний и отражается в росте итоговых метрик.

Калибровка вероятностей. Снижение ошибки калибровки (Таблица 5) является устойчивым эффектом в выбранном протоколе: на наборе ACDC метрика ESE уменьшается с 0,078 до 0,021. Метрика ESE вычислялась в поклассовой многоклассовой постановке (one-vs-rest) и агрегировалась по классам, что делает оценку чувствительной как к частым, так и к редким структурам. Проверка устойчивости вывода к способу биннинга (равноширинный против равнонаполненного) сохраняет качественные выводы, что снижает вероятность артефактов, связанных с неравномерным распределением предсказанных вероятностей. При этом улучшение калибровки следует интерпретировать как более согласованное соответствие «уверенность – точность» в пределах рассмотренных датасетов и протокола; оно не является клинической валидацией и не заменяет внешней проверки, анализа рисков и требований к безопасности в конкретных сценариях применения.

Карты неопределенности и выявление ошибок. Качество карт неопределенности оценивалось по способности ранжировать пиксели в задаче выявления ошибок

сегментации (Таблица 6), где положительный класс соответствует событию $\varepsilon(x) = I\{\hat{y}(x) \neq y(x)\}$. Метрика AUPRC интерпретируется как

$$AUPRC = \mathbb{P}(U(x^+) > U(x^-)),$$

где x^+ – случайный пиксель из множества ошибочных $\{\hat{y}(x) \neq y(x)\}$, x^- – случайный пиксель из множества корректных $\{\hat{y}(x) = y(x)\}$; пиксели класса void исключаются, фон учитывается. В наших экспериментах получено $AUPRC = 0,891$ (Таблица 6). При редком положительном классе (ошибки порядка $\approx 8\%$) более информативной метрикой является AUPRC; рост AUPRC ($0,284 \rightarrow 0,402$) указывает на улучшение качества ранжирования в выбранной постановке. Использование $U(x)$ для схем «отбраковки» (selective prediction) или маршрутизации на ручную проверку следует рассматривать как возможный прикладной сценарий: он требует отдельного выбора порогов, анализа чувствительности и оценки влияния на рабочий процесс с учетом баланса ложноположительных и ложноотрицательных срабатываний.

Согласованность дискретизации и выбор шага сетки. Теоретический результат о согласованности дискретизации энергии Дирихле (Теорема 1 при предположении о невырожденности и согласованности весов) согласуется с эмпирической проверкой на гладком синтетическом поле: отношение ошибок $e(h)/e(h/2)$ на последовательных сетках близко к 2 (Таблица 2), что соответствует первому порядку по h . В основных экспериментах базовый шаг $h = 1$ (соответствующий выбранному разрешению входных изображений) обеспечивает практический компромисс между точностью дискретизации регуляризатора и вычислительной эффективностью; дальнейшее уменьшение h снижает дискретизационную ошибку, но приводит к заметному росту потребления памяти и времени.

Вычислительные затраты. Согласно Таблице 7, предлагаемый метод сохраняет близкий к однопроходному режим инференса: время инференса возрастает с 22 до 26 мс на изображение при умеренном росте памяти ($7,6 \rightarrow 8,2$ GB). Для сравнения, MC-dropout требует T прогонов (в данной работе $T = 30$), а ансамбли – обучения и хранения МММ независимых моделей (здесь $M = 5$), что приводит к существенно большим затратам на инференс и/или обучение. Приведенные значения относятся к конкретной аппаратно-программной конфигурации (A100, batch size = 8) и служат ориентиром; достижимость режима, близкого к реальному времени, зависит от разрешения входных данных, оптимизации реализации и целевой платформы.

Ограничения и направления дальнейших исследований. Проведенная валидация имеет ряд ограничений. Во-первых, эксперименты выполнены в 2D-постановке; перенос на 3D требует отдельного исследования с учетом анизотропии вокселей и роста вычислительных затрат. Во-вторых, Теорема 1 устанавливает согласованность дискретизации для фиксированных гладких полей; распространение анализа на нейросетевые параметризованные поля требует дополнительных предположений о регулярности и, в более строгой постановке, рассмотрения Γ -сходимости семейства функционалов. В-третьих, в работе не изучалось поведение метода при сдвиге домена и на данных вне распределения обучения (междатасетное тестирование). В-четвертых, сравнение с архитектурно сильными базовыми решениями (например, nnU-Net v2) оставлено за рамками данной версии и необходимо для более строгого отделения эффекта функционала от архитектуры. Наконец, введенное разложение неопределенности требует эмпирической проверки ожидаемой динамики эпистемической компоненты при изменении объема обучающей выборки.

Заключение

Предложен вычислительный метод сегментации изображений на основе поля Дирихле, позволяющий получать вероятностное предсказание и карту неопределенности за один прямой проход нейросети (без многократных стохастических прогонов на инференсе). Эксперименты на датасетах ACDC, Synapse и CHAOS показали следующее.

1. Качество сегментации. На ACDC достигнут Dice 0,912 против 0,891 для базовой модели CE (mean \pm std по 10 запускам); различия статистически значимы по парному критерию Уилкоксона по 10 запускам и подтверждены на уровне пациентов ACDC ($n = 20$). Аналогичная тенденция улучшения наблюдается на Synapse и CHAOS (Таблица 4).

2. Калибровка вероятностей. Ошибка калибровки ECE в поклассовой многоклассовой постановке уменьшается на всех датасетах; на ACDC от 0,078 до 0,021. Замена способа биннинга (равноширинный/равнонаполненный) не меняет качественных выводов (Таблица 5).

3. Карты неопределенности. В задаче выявления ошибок сегментации на уровне пикселей получены AUPRC = 0,891 и AUPRC = 0,402, что превосходит энтропию softmax и дает значения, близкие по порядку к ансамблям при существенно меньших затратах на инференс (Таблицы 6 и 7).

4. Вычислительные затраты. По профилированию на GPU A100 (batch size = 8) накладные расходы составляют около +17 % по времени обучения и +18 % по времени инференса относительно CE; приведенные значения зависят от аппаратно-программной конфигурации и служат ориентиром (Таблицы 3 и 7).

С теоретической стороны установлена согласованность дискретизации энергии Дирихле при невырожденности и согласованности edge-aware весов (Теорема 1); эмпирическая проверка на гладком синтетическом поле подтверждает первый порядок по шагу сетки (Таблица 2). Абляционный анализ демонстрирует согласованный вклад компонент функционала: R_{KL} преимущественно улучшает калибровку, а $R_{D,h}[m]$ и $R_{D,h}[s]$ усиливают пространственную согласованность предсказаний и поля концентрации; анализ чувствительности не выявил заметной деградации метрик в исследованных диапазонах гиперпараметров.

К направлениям дальнейшей работы относятся перенос на 3D-постановку, сравнение с архитектурно сильными базовыми решениями (например, nnU-Net v2) на равных условиях, междатасетная проверка при сдвиге домена (OOD), а также эмпирическая валидация разложения неопределенности, включая ожидаемую динамику эпистемической компоненты при изменении объема данных.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Begoli E., Bhattacharya T., Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence*. 2019;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. *Information Fusion*. 2021;76:243–297. <https://doi.org/10.1016/j.inffus.2021.05.008>
3. Kompa B., Snoek J., Beam A.L. Second opinion needed: communicating uncertainty in medical machine learning. *npj Digital Medicine*. 2021;4. <https://doi.org/10.1038/s41746-020-00367-3>
4. Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: *Proceedings of the 33rd International Conference on*

- Machine Learning (ICML 2016)*, 19–24 June 2016, New York City, NY, USA. PMLR; 2016. P. 1050–1059.
5. Lakshminarayanan B., Pritzel A., Blundell Ch. Simple and scalable predictive uncertainty estimation using deep ensembles. In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 04–09 December 2017, Long Beach, CA, USA*. 2017. P. 6402–6413.
 6. Maddox W.J., Izmailov P., Garipov T., et al. A simple baseline for Bayesian uncertainty in deep learning. In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 08–14 December 2019, Vancouver, BC, Canada*. 2019. P. 13132–13143.
 7. Sensoy M., Kaplan L.M., Kandemir M. Evidential deep learning to quantify classification uncertainty. In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 03–08 December 2018, Montréal, Canada*. 2018. P. 3183–3193.
 8. Malinin A., Gales M.J.F. Predictive uncertainty estimation via prior networks. In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 03–08 December 2018, Montréal, Canada*. 2018. P. 7047–7058.
 9. Charpentier B., Zügner D., Günnemann S. *Posterior network: Uncertainty estimation without OOD samples via density-based pseudo-counts*. arXiv. URL: <https://arxiv.org/abs/2006.09239> [Accessed 15th January 2026].
 10. Jungo A., Reyes M. Assessing reliability and challenges of uncertainty estimations for medical image segmentation. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference: Proceedings: Part II, 13–17 October 2019, Shenzhen, China*. Cham: Springer; 2019. P. 48–56. https://doi.org/10.1007/978-3-030-32245-8_6
 11. Litjens G., Kooi Th., Bejnordi B.E., et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017;42:60–88. <https://doi.org/10.1016/j.media.2017.07.005>
 12. Kingma D.P., Ba J. *Adam: A method for stochastic optimization*. arXiv. URL: <https://arxiv.org/abs/1412.6980> [Accessed 15th January 2026].
 13. Chen T., Xu B., Zhang Ch., Guestrin C. *Training deep nets with sublinear memory cost*. arXiv. URL: <https://arxiv.org/abs/1604.06174> [Accessed 15th January 2026].
 14. Micikevicius P., Narang Sh., Alben J., et al. *Mixed precision training*. arXiv. URL: <https://arxiv.org/abs/1710.03740> [Accessed 18th January 2026].
 15. Bernard O., Lalande A., Zotti C., et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*. 2018;37(11):2514–2525. <https://doi.org/10.1109/TMI.2018.2837502>
 16. Kavur A.E., Gezer N.S., Barış M., et al. CHAOS Challenge – combined (CT-MR) healthy abdominal organ segmentation. *Medical Image Analysis*. 2021;69. <https://doi.org/10.1016/j.media.2020.101950>
 17. Guo Ch., Pleiss G., Sun Y., Weinberger K.Q. On calibration of modern neural networks. In: *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, 06–11 August 2017, Sydney, NSW, Australia. PMLR; 2017. P. 1321–1330.
 18. Ronneberger O., Fischer Ph., Brox Th. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference: Proceedings: Part III, 05–09 October 2015, Munich, Germany*. Cham: Springer; 2015. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

19. He K., Zhang X., Ren Sh., Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27–30 June 2016, Las Vegas, NV, USA*. IEEE; 2016. P. 770–778. <https://doi.org/10.1109/CVPR.2016.90>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Щетинин Евгений Юрьевич, доктор физико-математических наук, профессор, Севастопольский государственный университет, Севастополь, Российская Федерация.

e-mail: riviera-molto@mail.ru

ORCID: [0000-0003-3651-7629](https://orcid.org/0000-0003-3651-7629)

Evgeniy Yu. Shchetinin, Doctor of Physico-mathematical Sciences, Professor, Sevastopol State University, Sevastopol, the Russian Federation.

Шевчук Андрей Андреевич, аспирант, Севастопольский государственный университет, Севастополь, Российская Федерация.

e-mail: andreiluck11@yandex.ru

Andrey A. Shevchuk, Postgraduate, Sevastopol State University, Sevastopol, the Russian Federation.

Статья поступила в редакцию 29.01.2026; одобрена после рецензирования 10.03.2026; принята к публикации 18.03.2026.

The article was submitted 29.01.2026; approved after reviewing 10.03.2026; accepted for publication 18.03.2026.