

УДК 004.855.5

## УСТОЙЧИВАЯ К ШУМУ В РАЗМЕТКЕ СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ В ЗАДАЧЕ СЕГМЕНТАЦИИ ГЛИОМ НА МРТ ИЗОБРАЖЕНИЯХ

© 2020 г. Т. Н. Сапаров<sup>1,2</sup>, А. И. Курмуков<sup>1,3,\*</sup>, Б. Н. Широких<sup>1,2,4</sup>, С. В. Золотова<sup>5</sup>,  
А. В. Голанов<sup>5</sup>, М. Г. Беляев<sup>4</sup>, А. В. Далечина<sup>6</sup>

<sup>1</sup> Институт проблем передачи информации им. А.А. Харкевича РАН  
127994 Москва, Большой Каретный пер., д. 19, Россия

<sup>2</sup> Московский физико-технический институт (Государственный университет),  
141701 Московская обл., г. Долгопрудный, Институтский пер., 9, Россия

<sup>3</sup> Национальный исследовательский университет “Высшая школа экономики”,  
101000 Москва, Мясницкая ул., 20, Россия

<sup>4</sup> Сколковский институт науки и технологий  
143026 Москва, Большой бул., 30, Россия

<sup>5</sup> Федеральное государственное автономное учреждение  
“Национальный медицинский исследовательский центр нейрохирургии им. акад. Н.Н. Бурденко” МЗ РФ  
125047 Москва, 4-я Тверская-Ямская ул., 16, Россия

<sup>6</sup> Центр “Гамма-нож”, АО “Деловой центр нейрохирургии”  
125047 Москва, 1-й Тверской-Ямской пер., 13/5, Россия

\*E-mail: kurmukovai@gmail.com

Поступила в редакцию 05.06.2020 г.

После доработки 24.07.2020 г.

Принята к публикации 06.08.2020 г.

Сегментация медицинских изображений – одна из важнейших задач лучевой диагностики и терапии. Современные подходы к решению этой задачи основаны на глубоком обучении и показывают высокое качество при обучении на стандартизированных и специально собранных данных. Однако при работе с реальными клиническими изображениями ситуация кардинально меняется из-за принципиально более сложного устройства данных. В задаче сегментации опухолей головного мозга для планирования лучевой терапии размеры и интенсивности изображений существенно варьируются в зависимости от настроек аппарата магнитно-резонансной томографии; отмечается неоднозначность трактовки разными экспертами выявляемых на томограммах изменений; наконец, контуры мишени не всегда соответствуют изображению магнитно-резонансной томографии вследствие использования дополнительных модальностей при планировании облучения. В силу указанных причин сформированные выборки содержат большое количество шумных аннотаций. Мы предлагаем устойчивый алгоритм обучения, основанный на модификации традиционной архитектуры сверточной нейронной сети при помощи модуля для обучения весов, используемых в результирующей функции потерь (взвешенной перекрестной энтропии). Наша модель успешно борется с наличием шума в разметке и значительно уменьшает эффект высокой гетерогенности данных, повышая качество сегментации на 38%.

*Ключевые слова:* сегментация медицинских изображений, МРТ, глубокое обучение, глиобластома

DOI: 10.31857/S0235009220040071

### ВВЕДЕНИЕ

Сегментация медицинских изображений является важнейшим этапом анализа данных медицинской визуализации, позволяющим выделить границы интересующей области (патологического очага или здоровых тканей) на изображении и получить ее количественные характеристики. Эта информация может быть использована как для постановки диагноза, так и для определения точных границ опухоли перед подготовкой к хирур-

гическому или лучевому лечению. Создание контура патологического очага (оконтуривание) является неотъемлемой частью планирования лучевой терапии и оценки результатов лечения. Высокая детальность медицинских изображений делает ручное оконтуривание крайне трудоемкой процедурой, требующей существенных временных затрат.

Задача автоматической сегментации медицинских изображений – одна из перспективных об-

ластей применения технологий на основе глубоких сверточных сетей (Hesamian et al., 2019). На сегодняшний день наиболее активно развивающееся направление — это сегментация данных магнитно-резонансной томографии (МРТ). Сверточные нейронные сети (СНС) успешно применяются для сегментации различных органов и тканей на МРТ изображениях: сегментации очагов рассеянного склероза, опухолей головного мозга, определения контуров сердца, печени, межпозвоночных дисков и другое (García-Lorenzo et al., 2013; Dvorak, Menze, 2015; Chen et al., 2020; Zeng et al., 2019; Kim et al., 2018). Представленные в литературе алгоритмы сегментации изображений построены, как правило, на модельных задачах при условии стандартизации протоколов сканирования и данных разметки. Адаптация таких моделей к реальным клиническим данным весьма затруднительна (Wang et al., 2019; Sahiner et al., 2019; Kelly et al., 2019).

Проблема “шума” в разметке при использовании алгоритмов обучения с учителем давно известна в литературе (Frénaу, Verleysen, 2013; Algan, Ulusoy, 2020), в частности, в задачах классификации (Karimi et al., 2020) и сегментации медицинских изображений (Tajbakhsh et al., 2020). Последние исследования показывают, что современные глубокие нейронные сети способны запоминать шумную разметку на этапе обучения (Zhang et al., 2017a), что приводит к переобучению, поэтому необходимо разрабатывать алгоритмы, устойчивые к шуму в разметке. К источникам “шума” относятся следующие факторы: недостаток данных, предоставленных эксперту в процессе разметки; субъективность экспертной разметки (Vinod et al., 2016; Growcott et al., 2020) (inter rater reliability); ошибки разметки. Здесь и далее мы употребляем слово “шум” в смысле шума в целевой переменной: разметке целевых областей, например патологических очагов. Изучение наличия шума в самих МРТ изображениях лежит за рамками данной работы.

Для борьбы с шумом в разметке в задачах с использованием глубоких нейронных сетей было разработано множество методов. Для задач классификации изображений авторы работы (Tanaka et al., 2018) предлагают “исправлять” разметку в процессе обучения: нейронная сеть обучается на существующей шумной разметке, затем она уточняется предсказаниями сети, и на новой итерации обучения используются уже уточненные метки. Применительно к сегментации медицинских изображений авторы используют идеи генеративно-сопоставительных нейронных сетей (generative adversarial network, GAN), предлагая архитектуры, в которых вместо генератора используется сегментационная СНС, а задача дискриминатора заключается в отделении точной (предсказанной) сегментации от неточной (Zhang et al., 2017 б) или

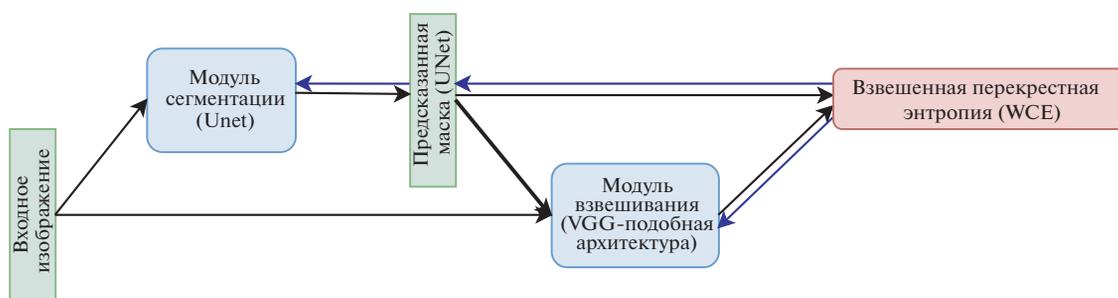
экспертной разметки от предсказания сети (Nie et al., 2018).

В ряде работ предлагаются различные способы взвешивания изображений с шумом и без шума в разметке, например, авторы (Mirikharaji et al., 2019) обучали нейросеть попеременно на изображениях с шумом в разметке и без. Изображения с шумом учитываются в результирующей функции потерь с весами таким образом, чтобы изменение параметров нейросети не ухудшало предсказаний на изображениях без шума в разметке. Авторы (Xue et al., 2019) решают задачу классификации меланомы и предлагают отбирать изображения в процессе обучения: на этапе обучения нейросети, после получения предсказаний наблюдения сортируются по убыванию уверенности сети и для обновления весов используют только те примеры, для которых эта уверенность превышает некоторый, заранее заданный, порог. В работе (Zhu et al., 2019) предлагается использовать взвешенную перекрестную энтропию с автоматическим подбором весов для выявления примеров с ошибочной разметкой и снижения их влияния на обучение.

В нашей работе мы развиваем идею адаптивного перевзвешивания объектов для построения устойчивого алгоритма сегментации. Для этого используется дополнительный модуль СНС (рис. 1, Модуль взвешивания), который принимает на вход исходное изображение и предсказанную модулем сегментации СНС (рис. 1. Модуль сегментации) карту вероятностей. Задача модуля взвешивания оценить “вес” изображения, чтобы он убывал с ростом шума в разметке. Затем эти веса комбинируются с выходом модуля сегментации и подаются в функцию потерь. Оба модуля являются компонентами одной СНС и обучаются совместно (для обучения параметров модуля взвешивания не используется никакая дополнительная разметка).

Ключевые отличия нашей работы от работы (Zhu et al., 2019) состоят в следующем.

- Во-первых, в работе (Zhu et al., 2019) для оценки веса используется не карта вероятностей полученных СНС, а экспертная разметка. Это отличие кардинально меняет интерпретацию получаемых весов и повышает риски переобучения.
- Во-вторых, мы предлагаем интерпретируемую регуляризацию обучаемых весов. В зависимости от значения параметра регуляризации метод может действовать в целом диапазоне стратегий обучения, начиная от обычной сегментации без перевзвешивания (равные веса), и заканчивая занулением всех весов, кроме одного.
- В-третьих, вместо простой задачи сегментации легких с синтетическим шумом в целевых контурах мы используем реальные данные с су-



**Рис. 1.** Схема предложенной СНС. В экспериментах без взвешивания примеров используется только сегментационная нейросеть (Модуль сегментации). Блоки зеленого цвета – входные и выходные тензоры, блоки синего цвета – нейронные сети, блок красного цвета – функция потерь. Стрелки черного цвета обозначают прямой проход при вычислениях, синего – обратный проход (вычисление градиента). Модулю взвешивания предсказанная маска подается как независимый тензор.

шественно более сложной задачей сегментации опухолей головного мозга на МРТ изображениях.

## ОПИСАНИЕ ДАННЫХ

МРТ изображение – трехмерный массив, элементы которого называются вокселями (volume pixel). Эти изображения могут быть получены в различных ортогональных плоскостях: коронарной, сагиттальной, аксиальной. Плоскость, в которой изображение имеет наибольшее разрешение, будем называть главной. Роль аппаратной вариабельности можно оценить на примере рис. 2 и 3. Для получения изображений, представленных на рисунках, использовалась одна импульсная МРТ последовательность T2 FLAIR (Fluid Attenuated Inversion), но разные томографы, что является причиной кардинального отличия в распределении интенсивностей пикселей. Разница в толщине двумерных срезов, расстоянии между срезами, а также разрешении главной проекции приводит к расхождениям в физическом объеме вокселей.

Для создания модели автоматической сегментации были использованы обезличенные трехмерные МРТ изображения пациентов со злокачественной внутримозговой опухолью – глиобластомой. Выборка включала 185 пациентов, которые получали лучевое лечение ложа удаленной опухоли с 2014 по 2019 гг. на медицинских линейных ускорителях Novalis (BrainLab) и Truebeam (Varian) в ФГАУ НМИЦ нейрохирургии им. акад. Н.Н. Бурденко: 98 изображений были представлены аксиальными срезами, 87 – сагиттальными. Толщина срезов от 0.8 до 5 мм, расстояние между срезами от 1 до 6.5 мм (рис. 4). Данные получены более чем с 20 различных моделей томографов.

Контуры мишеней были созданы в планирующих системах iPlan (BrainLab) и Eclipse (Varian). В разметке участвовали 10 врачей – радиотерапевтов с опытом работы от 5 до 15 лет (медиана 11 лет)

в отделении радиохирургии и радиотерапии ФГАУ НМИЦ нейрохирургии им. акад. Н.Н. Бурденко. Во время процедуры планирования контур ложа глиобластомы GTV (gross tumor volume) определялся в большинстве случаев по МРТ в режиме FLAIR (в объем GTV включалась вся зона гиперинтенсивного сигнала в режиме FLAIR) и мог корректироваться с учетом дополнительных последовательностей.

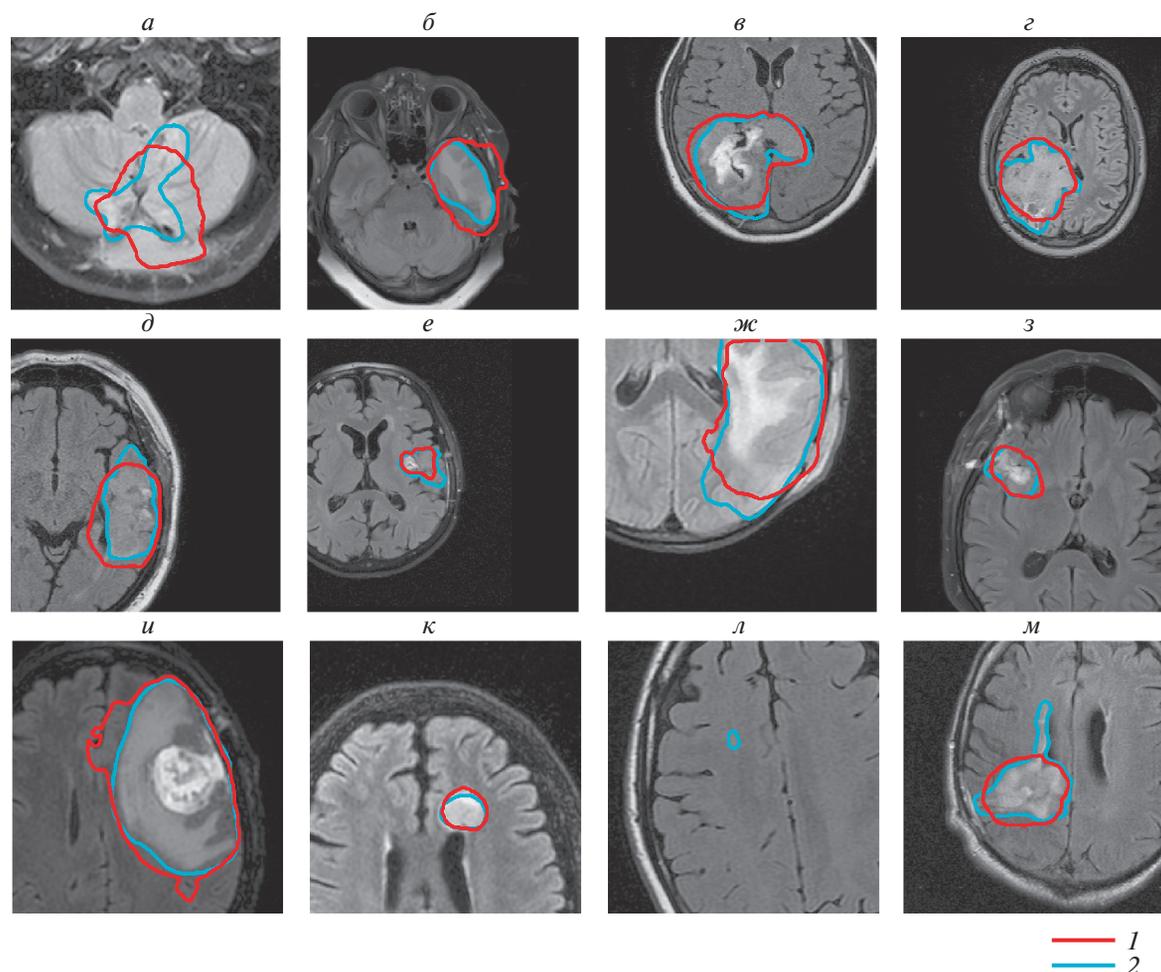
### *Источники шума в разметке глиобластом*

Решение задачи автоматической сегментации ложа удаленной глиобластомы для планирования лучевой терапии осложняется следующими факторами.

- Неоднозначность определения границ опухоли (в особенности после удаления) (Zhao et al., 2016).
- Отсутствие экспертной согласованности при сегментации мишени (Vinod et al., 2016; Growcott et al. 2020; Sandström et al. 2018).
- Экспертная разметка не совпадает с зоной гиперинтенсивного сигнала в режиме FLAIR.

Последняя особенность связана с тем, что контур мишени, определенный по FLAIR, как правило, наиболее объективно отражает распространенность инвазии глиобластомы, однако для уточнения границ опухоли используются несколько МРТ последовательностей (T1, T2, T1 с контрастным усилением) и другие дополнительные модальности (позитронно-эмиссионная томография, ПЭТ). Поэтому в ряде случаев контур мишени может существенно отличаться от зоны гиперинтенсивного сигнала по FLAIR (рис. 3), где на  $a$  область менее интенсивного сигнала также включена в основной контур мишени, а на  $z$  экспертный контур не захватывает всю область гиперинтенсивного сигнала по FLAIR.

Перечисленные проблемы значительно осложняют как **обучение** СНС, так и **оценку качества работы** итогового алгоритма сегментации.



**Рис. 2.** Изображения, на которых экспертная разметка не отличается от зоны гиперинтенсивного сигнала в режиме FLAIR. Предсказания алгоритма сегментации (1 – красный контур); экспертная разметка (2 – синий контур).

## МЕТОДЫ ИССЛЕДОВАНИЯ

Зафиксируем следующие обозначения:  $X_i$  – патч (прямоугольная часть изображения) размера  $M \times M$ ,<sup>1</sup> подаваемая на вход сети,  $Y_i$  – соответствующая экспертная разметка (матрица размера  $M \times M$ , состоящая из нулей и единиц; единица означает, что соответствующий пиксель патча  $X_i$  содержит опухоль),  $\hat{Y}_i$  – вероятностная карта (также имеет размер  $M \times M$ ), получаемая как выход сегментационной сети для изображения  $X_i$ ,  $b$  – размер батча. Каждый батч состоит из  $b$  патчей размера  $M \times M$  (патчи выбираются случайным образом из объектов тренировочной выборки).

В качестве базовой модели сегментации мы используем UNet (Ronneberger et al., 2015) (рис. 5). В связи с большой вариабельностью толщин дву-

мерных срезов и расстояниями между срезами мы решили использовать версию Unet с двумерными свертками и учиться на отдельных двумерных срезах, а не на трехмерных изображениях. На вход сети подается батч из частей изображения размера  $240 \times 240$ . Батч состоит из 32 таких изображений. В качестве аугментаций были использованы повороты и отражения относительно горизонтальной и вертикальной осей, а также гамма преобразования интенсивностей. Аугментации проводились непосредственно во время обучения.

Ключевой компонент нашего метода – это использование взвешенной функции потерь с адаптивным выбором весов для каждого изображения на каждой итерации обучения. Такой подход позволяет уменьшить вклад в общую функцию потерь тех изображений, разметка которых содержит шум. Мы предлагаем новый метод для оценки весов наблюдений при помощи дополнительного модуля СНС, а также сравниваем его с популярным подходом, в котором вес является некоторой функ-

<sup>1</sup> В процессе обучения значение  $M$  приняли равным 240. Размер был выбран как компромисс между размером патча и количеством патчей в одном батче.

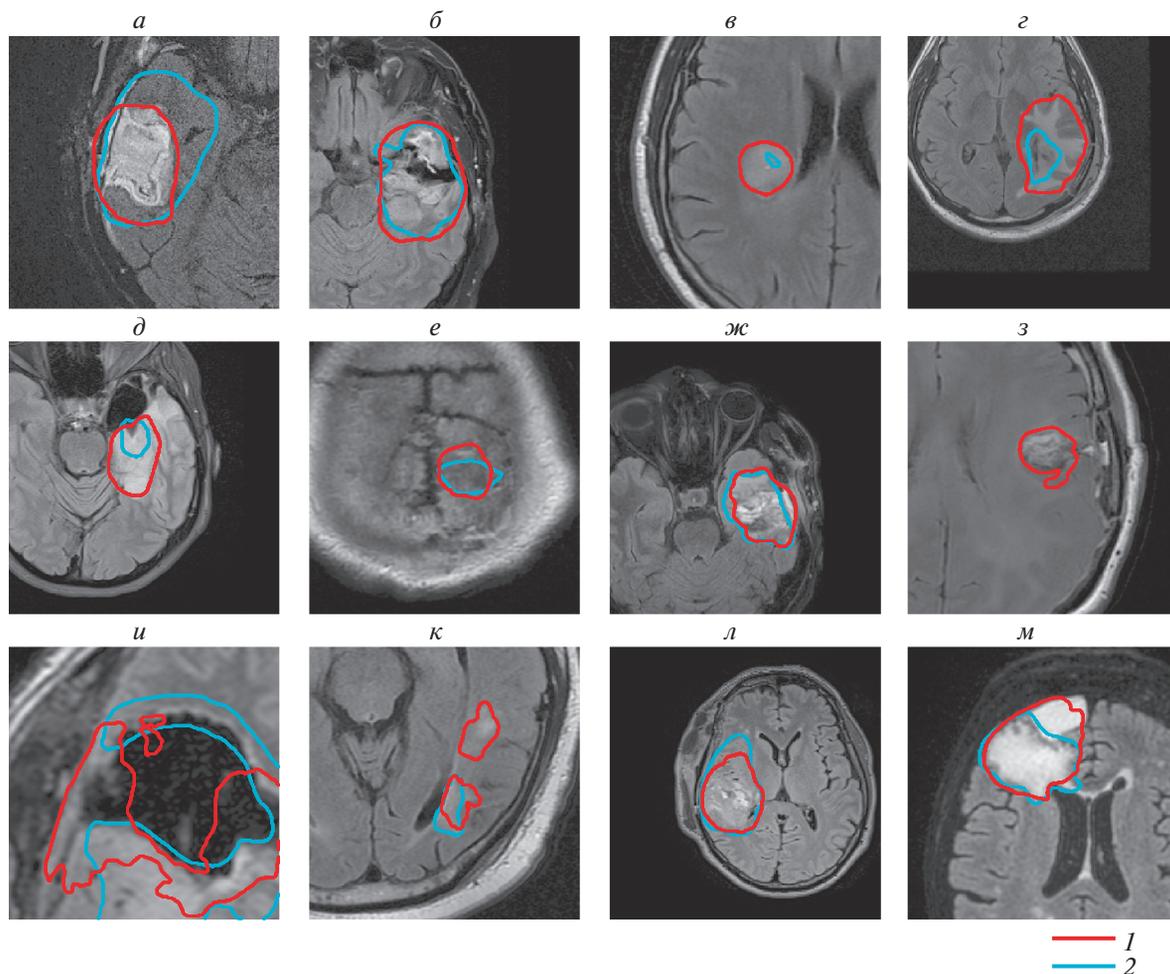


Рис. 3. Изображения, на которых экспертная разметка отличается от зоны гиперинтенсивного сигнала в режиме FLAIR. Предсказания алгоритма сегментации (1 – красный контур); экспертная разметка (2 – синий контур).

цией от значения функции ошибки. В обоих случаях нейронная сеть обучается первые 50 эпох без дополнительного взвешивания для предварительной настройки основной сегментационной сети.

**Функция потерь**

В качестве функции потерь используется взвешенная перекрестная энтропия:

$$\begin{aligned}
 WCE(\hat{Y}, Y) &= \sum_{i=1}^b w_i BCE(\hat{Y}_i, Y_i), \\
 BCE(\hat{Y}_i, Y_i) &= \\
 &= \frac{1}{M^2} \sum_{j,k=1}^M Y_i^{jk} \log(\hat{Y}_i^{jk}) + (1 - Y_i^{jk}) \log(1 - \hat{Y}_i^{jk}),
 \end{aligned}$$

где  $Y$  – батч экспертных разметок, а  $\hat{Y}$  – батч соответствующих им предсказанных вероятностных карт,  $Y_i^{jk}$  и  $\hat{Y}_i^{jk}$  – соответствующие пиксели  $Y$  и  $\hat{Y}$ , а для подбора весов  $w_i$  используются две различные стратегии.

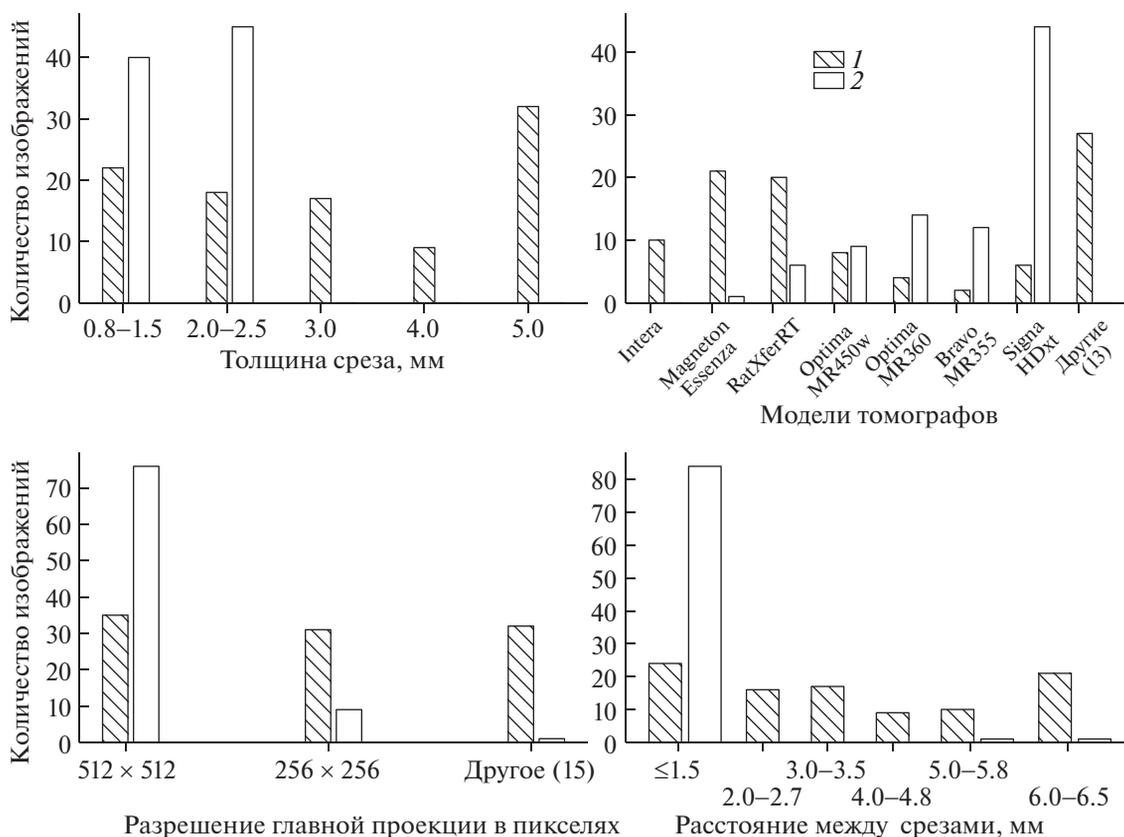
**Обучаемые веса**

Первый подход заключается в использовании обучаемых весов. Такие веса являются выходами вспомогательного модуля сверточной сети (рис. 6). Этот модуль принимает на вход одновременно  $X_i$  и  $\hat{Y}_i$ , объединенные в двухканальное изображение. Выходом является вещественное число ( $l_i$ ), которое преобразуется непосредственно в вес применением функции softmax:

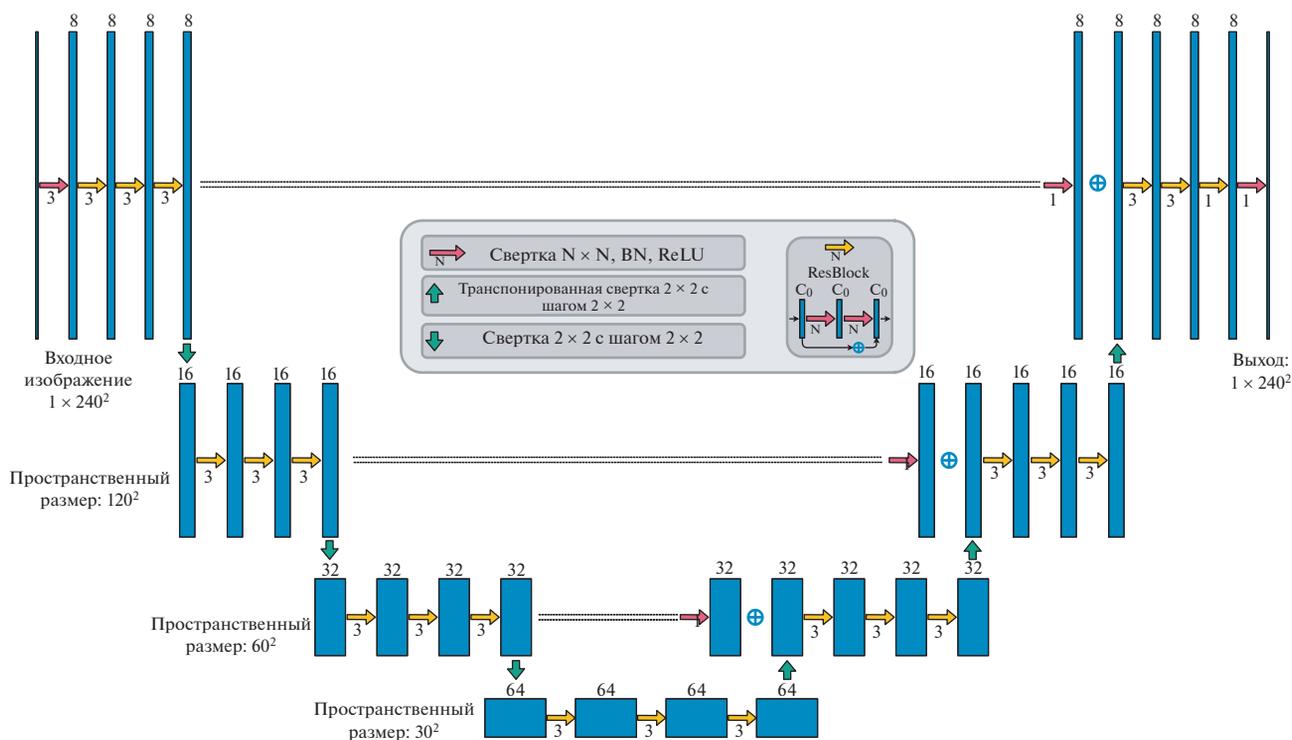
$$w_i = \text{softmax}(l_i) = \frac{\exp(l_i)}{\sum_{j=1}^b \exp(l_j)}.$$

Для того чтобы избежать ситуации, когда одно из  $l_i$  является много больше других по значению, вследствие чего распределение на выходе из softmax оказывается вырожденным (а эффективный размер батча становится единицей), в функцию потерь было добавлено регуляризирующее слагаемое:

$$WCE(\hat{Y}, Y) = \sum_{i=1}^b w_i BCE(\hat{Y}_i, Y_i) + \alpha \sum_{i=1}^b w_i^2.$$



**Рис. 4.** Вариабельность используемых изображений по толщинам двумерных срезов, моделям томографов, разрешению изображения в главной (двумерной) проекции и расстоянию между срезами. 1 – для аксиальной проекции; 2 – для сагиттальной проекции.



**Рис. 5.** Архитектура модуля сегментации. Синие блоки – тензоры. Число над синим блоком обозначает число каналов.

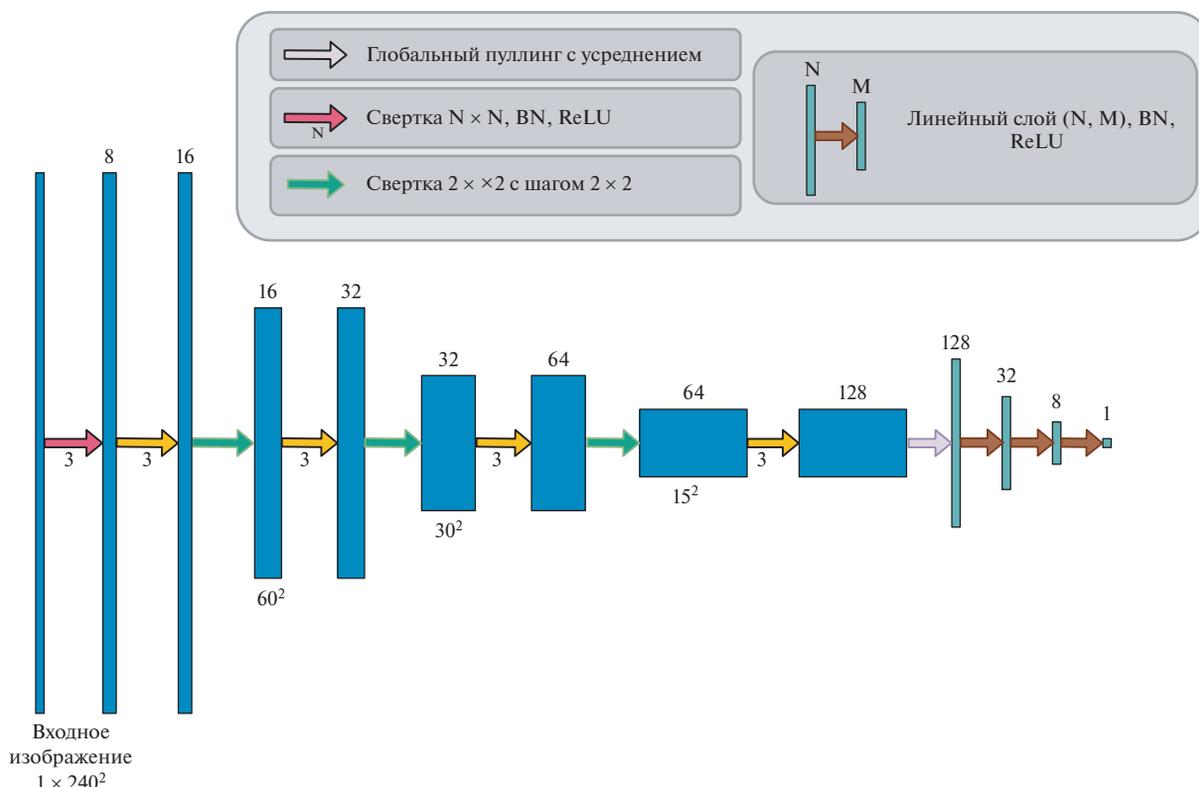


Рис. 6. Архитектура модуля взвешивания. Синие блоки – тензоры, зеленые блоки – векторы в полносвязной части сети.

Параметр  $\alpha$  на обучении был равен 0.1. Изменение коэффициента регуляризации  $\alpha$  позволяет балансировать между конфигурацией, в которой вес одного изображения оказывается значительно больше веса всех остальных (в результате батч-градиентный спуск вырождается в стохастический градиентный спуск), и конфигурацией, в которой используется одинаковый вес для всех изображений (и взвешенная перекрестная энтропия превращается в обычную). Дополнительный модуль СНС, выходом которого являются веса изображений ( $w_i$ ), используется **только** на этапе обучения и обучается совместно с модулем сегментации, путем минимизирования функции потерь  $WCE$ .

*Веса, линейно зависящие от функции потерь*

В данном случае предлагается использовать веса, которые будут линейно зависеть от значения функции потерь. Зависимость определяется таким образом, что в батче для изображения с наименьшим значением функции потерь вес будет наибольшим, и наоборот. Веса рассчитываются по следующей формуле:

$$w_i = \frac{BCE_{\max} - BCE(\hat{Y}_i, Y_i)}{BCE_{\max} - BCE_{\min}}$$

где  $BCE_{\max}$  – это максимальное значение функции потерь в батче, а  $BCE_{\min}$  – соответственно, минимальное. После этого каждый из рассчитанных весов делится на сумму всех весов в батче для того, чтобы нормировать их к единице.

**МЕТРИКА КАЧЕСТВА**

Для оценки качества сегментации использовался индекс Дайса  $\times 100$ :

$$Dice(Y_i, \hat{Y}_i) = \frac{2 \sum_{i,j=1}^M Y_i^{jk} \hat{Y}_i^{jk}}{\sum_{i,j=1}^M Y_i^{jk} + \sum_{i,j=1}^M \hat{Y}_i^{jk}} \times 100.$$

Пороговое значение, по которому вероятностная карта преобразовывалась в бинарную маску опухоли, выбирается адаптивно (индивидуально на каждом изображении) при помощи алгоритма Отсу (Otsu, 1979). В предварительных экспериментах такой способ подбора порога показал качество не хуже, чем подбор оптимального порога на отложенной выборке (при этом значительно лучше, чем при использовании постоянного порога 0.5).

*Кросс-валидация*

Обучение проводилось с использованием перекрестного контроля на пяти подвыборках. Экс-

**Таблица 1.** Точность сегментации измерена в терминах меры Дайса. Среднее значение (стандартное отклонение)

Эксперимент	Аксиальная проекция	Сагиттальная проекция
Базовая модель (без взвешивания)	47 (23)	71 (14)
Линейные веса	61 (19)	74 (13)
Обучаемые веса	65 (17)	71 (13)

пертная разметка содержит контуры, несогласованные с зоной гиперинтенсивного сигнала на FLAIR, что не позволяет использовать всю выборку для корректного сравнения работы алгоритмов.<sup>2</sup> Поэтому для сравнения работы моделей до начала обучения были вручную отобраны 45 снимков в аксиальной (основной) проекции и 57 – в сагиттальной, с разметкой, визуально совпадающей с изменениями по FLAIR. В каждую из подвыборок на кросс-валидации стратифицировано попадали как заранее отобраные снимки, так и оставшиеся снимки с шумом в разметке. Таким образом, для обучения использовались все данные, а для подсчета метрики Дайса на тестовой выборке только данные без шума в экспертной разметке.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Итоговые результаты представлены в табл. 1. Для анализа эффективности модуля автоматического взвешивания было решено сравнить качество сегментации при использовании взвешивания и без него. В качестве базовой модели использовалась архитектура Unet (рис. 5). После 50 эпох обучения (средний индекс Дайса после первых 50 эпох без дополнительного взвешивания – 0.36 для тестовой выборки с аксиальной главной плоскостью и 0.45 – для сагиттальной) оно было независимо продолжено без взвешивания, с применением линейного взвешивания и с применением обучаемых весов (табл. 1). Мы наблюдаем значительный рост качества сегментации в терминах индекса Дайса (0.65 индекс Дайса) для модели, построенной на аксиальных снимках (данные с высокой вариабельностью томографов и размеров изобра-

<sup>2</sup> Для оценки качества используется метрика, сравнивающая предсказание алгоритма и экспертную разметку, но если в разметке присутствует шум, то возможны ошибки первого рода: предсказание алгоритма точно, но не совпадает с экспертной разметкой, и мы наказываем алгоритм за неверное предсказание, которое на самом деле верно. Это создает проблему как на этапе обучения (цель нашей работы – разрешить эту проблему), так и на этапе оценки качества работы алгоритма. Для того чтобы корректно оценить работу алгоритма в тестовой выборке, использовались только те примеры, в которых экспертные контуры совпадают с зоной гиперинтенсивности на FLAIR.

жений) как относительно базовой модели без взвешивания (0.47 индекс Дайса), так и относительно модели со взвешиванием на основе значений функции потерь (0.61 индекс Дайса). Для модели, построенной на сагиттальных снимках, наблюдается незначительный прирост качества сегментации, при этом взвешивание на основе значений функции потерь оказалось даже несколько лучше обучаемых весов (0.74 против 0.71 индекс Дайса). Мы предполагаем, что эффект предложенного взвешивания на сагиттальных снимках оказался ниже, потому что они содержат меньше “шума” в разметке: 57 (~66%) из 87 сагиттальных содержат разметку, совпадающую с областью гиперинтенсивности на FLAIR, тогда как среди аксиальных снимков только 45 (~46%) из 98. Кроме этого, сагиттальные снимки значительно более гомогенны (по сравнению с аксиальными) по таким параметрам, как разрешение, толщина среза и модель томографа (рис. 4). Воздействие этого фактора должно стать предметом дальнейших исследований.

На рис. 7 изображено изменение весов, присваиваемых 24 патчам (рис. 2, *a–l* и рис. 3, *a–l*) модулем автоматического взвешивания в процессе обучения (для построения этого графика были сохранены веса нейросети на разных этапах обучения). Можно видеть, что изображениям, на которых разметка совпадает с областью патологических изменений, в среднем присваиваются большие веса, т.е. они сильнее влияют на параметры нейросети.

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

За последние 10 лет в 100 000 медицинских центрах по всему миру приблизительно у 140 млн. человек были диагностированы онкологические заболевания. В зависимости от медицинского учреждения на каждого пациента приходится от 0.1–10 Гб различных данных. Общий объем данных оценивается в 14–1400 петабайт (Lutsberg et al., 2017). Большую часть этого массива составляют данные медицинской визуализации, используемые для диагностики, планирования и оценки результатов лечения. Анализ этих данных при помощи глубокого обучения является перспективным методом извлечения необходимой информации для построения систем принятия решений и автоматизации рутинных процессов в онкологии.

Тем не менее разработка подобных алгоритмов на основе ретроспективных данных сопряжена с рядом трудностей. Несмотря на огромный массив гетерогенной информации, накопленной в клинике, при построении моделей приходится использовать ограниченный набор данных (Sahiner et al., 2019). Это связано как с особенностями экспорта из планирующих систем и рентгенологических архивов, так и с особенностями самих данных. В боль-

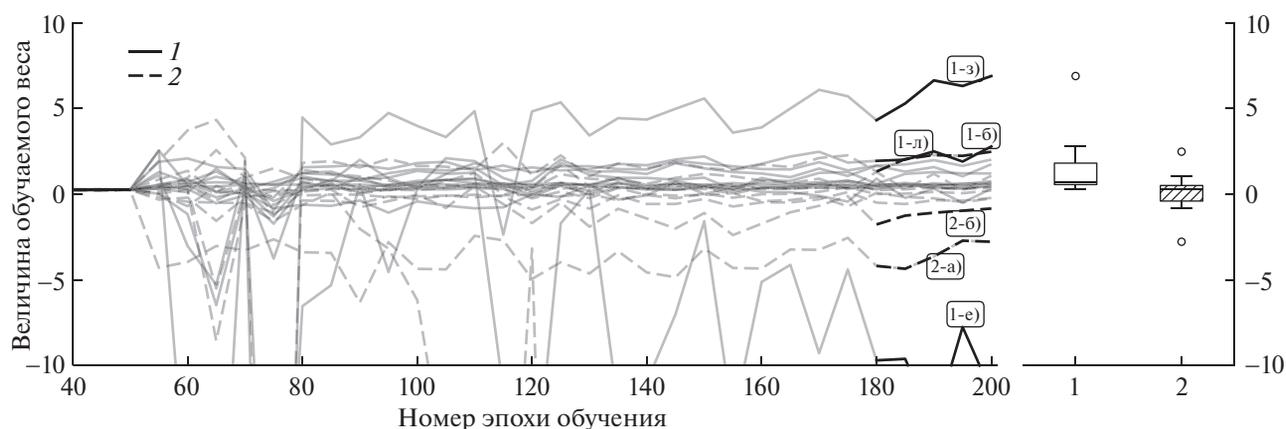


Рис. 7. Величина обучаемых весов для 24 патчей. Изменение веса изображения в процессе обучения (слева). Распределения весов с разметкой, визуально совпадающей с областью поражения (1) и не совпадающей (2) (справа).

шинстве коммерческих системах планирования, использующихся в отделениях лучевой терапии, не предусмотрена возможность автоматического экспорта планов в формате DICOM RT. Поэтому сбор ретроспективных данных является крайне трудоемким процессом, выполняющимся вручную. Так, например, в представленной работе при решении задачи сегментации глиобластомы экспорт файлов из планирующей системы Eclipse занимал в среднем 5 мин на пациента. Экспорт данных для нескольких сотен пациентов требует существенных временных ресурсов.

Вариабельность разметки является одной из самых серьезных проблем при анализе медицинских изображений. В литературе широко освещена тема отсутствия единого стандарта оконтуривания для многих патологических образований (Vinod et al., 2016; Growcott et al., 2020; Sandström et al., 2018). Особенно актуальна эта проблема для сегментации опухолей после их частичного удаления. Определение оптимального объема облучения опухолей головного мозга остается крайне спорным вопросом, ответ на который зависит от внутренних протоколов исследовательских групп и нейроонкологических клиник. В работе (Zhao et al., 2016) показана высокая субъективность при определении контура глиобластомы даже внутри одного центра. Авторы отмечают, что 61% клиницистов выбирали в качестве объема для облучения ложе удаленной опухоли и область повышенного накопления контраста с краевым захватом, 33% участников исследования включали в объем облучения область отека в дополнении к послеоперационной полости и контрастирующей части опухоли.

Результаты нашей работы также свидетельствуют о сильной вариабельности контуров глиобластомы на послеоперационных изображениях. К примеру, на рис. 3, г представлен аксиальный срез МРТ исследования пациента с глиобластомой, на котором изначальная разметка не соответствует области изменения по FLAIR. При планировании

облучения GTV определялся по изображениям T1 с контрастным усилением, к которому в дальнейшем добавлялся краевой захват, включающий область гиперинтенсивного сигнала по FLAIR. При расположении мишени вблизи органов риска врач мог уменьшать краевой захват. Также контур мишени не включает всю зону гиперинтенсивного сигнала по FLAIR, если рядом с опухолью находятся изменения неопухолевого генеза (ишемия, воспаление).

В этом исследовании впервые представлены результаты работы модели сегментации глиобластомы после частичного удаления, построенной на ретроспективных данных без предварительного отбора случаев с согласованной разметкой и устранения вариабельности в протоколах сканирования. Предложенный метод модификации традиционной архитектуры сверточной нейронной сети позволяет добиться увеличения метрики Дайс с 47 до 65 на аксиальных изображениях. Стоит отметить, что результаты работ по сегментации, представленные в литературе, демонстрируют более высокие показатели метрик качества (например, на послеоперационных изображениях достигается Дайс более 80) (Bakas et al., 2018). Однако эти результаты получены на выборках данных большего объема, собранных по единому протоколу, в том числе в рамках проведения одноцентровых/мультицентровых исследований (Kickingreder et al., 2019). Существенным ограничением этих работ является сегментация МРТ изображений, полученных на одном томографе или при использовании одинаковых протоколов сканирования, что лимитирует адаптацию разработанных алгоритмов для клинической практики. Предложенная нами нейросетевая архитектура построена на основании МРТ исследований, выполненных на 20 моделях томографов с различными параметрами сканирования. При этом предложенный в данной работе метод позволяет повысить качество сегментации на ретроспектив-

ных данных, несмотря на вариабельность МРТ томографов и протоколов сканирования.

### ЗАКЛЮЧЕНИЕ

Построение моделей автоматической сегментации МРТ изображений с использованием реальных данных, накопленных в клиниках, связано с тремя проблемами: высокий уровень вариабельности экспертной разметки, несоответствие разметки изображению и большое разнообразие в протоколах сканирования. Каждая из них вносит существенную неопределенность в обучение ней-

ронной сети, препятствуя успешному применению технологий глубокого обучения для ретроспективных данных.

В этой работе мы продемонстрировали, как модификация традиционной архитектуры сверточной нейросети, а именно, добавление модуля обучаемых весов и использование их в результирующей функции потерь, позволяют эффективно бороться с проблемами, вызванными гетерогенностью данных и шумом в разметке.

Исследование было поддержано РФФИ (грант № 18-29-01054).

## A convolutional neural network, robust to label noise for glioma segmentation

T. N. Saparova<sup>a,b</sup>, A. I. Kurmukov<sup>a,c,#</sup>, B. N. Shirokih<sup>a,b,d</sup>, S. V. Zolotova<sup>e</sup>, A. V. Golovanov<sup>e</sup>,  
M. G. Belyaev<sup>d</sup>, and A. V. Dalechina<sup>f</sup>

<sup>a</sup> Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute),  
127994 Moscow, Bol'shoy Karetny Pereulok, 19, Russian

<sup>b</sup> Moscow Institute of Physics and Technology, 141701 Dolgoprudny,  
Moscow Region, Institutskiy Pereulok, 9, Russian

<sup>c</sup> Higher School of Economics – National Research University,  
101000 Moscow, Myasnitskaya Ulitsa, 20, Russian

<sup>d</sup> Skolkovo Institute of Science and Technology,  
143026 Moscow, Bol'shoy Bul'var, 30, Russian

<sup>e</sup> Federal State Autonomous Institution "N.N. Burdenko National Medical Research Center of Neurosurgery"  
of the Ministry of Health of the Russian Federation,  
125047 Moscow, 4-Ya Tverskaya-Yamskaya Ulitsa, 16, Russian

<sup>f</sup> Moscow Gamma Knife Center, Joint Stock Company "Neurosurgery Business Center",  
125047 Moscow, 1-Y Tverskoy-Yamskoy Pereulok, 13/5, Russian

#E-mail: kurmukovai@gmail.com

Medical image segmentation is one of the most important tasks in radiotherapy. Recently, deep neural networks have achieved great success in solving a wide variety of medical image segmentation tasks using standardized or/and hardly preprocessed data. However, with typical clinical data, the situation drastically changes because of data complexity. For example, in radiotherapy tasks, data are much more variable in sizes and pixel intensities depending on a particular MRI machine. This heterogeneity leads to high intra- and inter-rater variability in tumor delineation. Finally, tumor contours might not coincide with pathological regions on the MRI image due to complex delineation procedures. Thus, the annotations in the existing datasets are relatively noisy. We propose a robust learning procedure based on the modern Convolutional Neural Network architecture. Our extension uses a separate module with learnable parameters, which outputs weights for the weighted cross-entropy. In our experiments, we achieved an impressive 38% boost in the Dice score using the data with high variability and target noise.

*Key words:* medical image segmentation, MRI, deep learning, high grade glioma

### REFERENCES

- Algan G., Ulusoy I. Image Classification with Deep Learning in the Presence of Noisy Labels: A Survey. <https://arxiv.org/pdf/1912.05170.pdf> (accessed 01.06.2020)
- Bakas S., Reyes M., Jakab A., Bauer S., Rempfler M. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. <https://arxiv.org/pdf/1811.02629.pdf> (accessed 05.11.2018)
- Chen C., Qin C., Qui H., Tarroni G., Duan J. Deep learning for cardiac image segmentation: A review. *Frontiers in Cardiovascular Medicine*. 2020. V. 7. P. 25. <https://doi.org/10.3389/fcvm.2020.00025>
- Dvorak P., Menze B. Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. *Proceeding of the multimodal brain tumor image segmentation challenge*. 2015. P. 13–24.
- Frénay B., Verleysen M. Classification in the presence of label noise: a survey. *IEEE transactions on neural networks and learning systems*. 2013. V. 25(5). P. 845–869. <https://doi.org/10.1109/tnnls.2013.2292894>
- García-Lorenzo D., Francis S., Narayanan S., Arnold D.L., Collins D.L. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical image analysis*. 2013. V. 17(1). P. 1–18. <https://doi.org/10.1016/j.media.2012.09.004>
- Growcott S., Dembrey T., Patel R., Eaton D., Cameron A. Inter-Observer Variability in Target Volume Delinea-

- tions of Benign and Metastatic Brain Tumours for Stereotactic Radiosurgery: Results of a National Quality Assurance Programme. *Clinical Oncology*. 2020. V. 32 (1). P. 13–25.  
<https://doi.org/10.1016/j.clon.2019.06.015>
- Hesamian M.H., Jia W., He X., Kennedy P. Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of digital imaging*. 2019. V. 32 (4). P. 582–596.  
<https://doi.org/10.1007/s10278-019-00227-x>
- Karimi D., Dou H., Warfield S.K., Gholipour A. Deep learning with noisy labels: exploring techniques and remedies in medical image analysis.  
<https://arxiv.org/pdf/1912.02911.pdf> (accessed 01.06.2020)
- Kelly C.J., Karthikesalingam A., Suleyman M., Corrado G., King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*. 2019. V. 17 (1). P. 195.  
<https://doi.org/10.1186/s12916-019-1426-2>
- Kickingereder P., Isensee F., Tursunova I., Petersen J., Neuberger U. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *The Lancet Oncology*. 2019. V. 20 (5). P. 728–740.  
[https://doi.org/10.1016/s1470-2045\(19\)30098-1](https://doi.org/10.1016/s1470-2045(19)30098-1)
- Kim S., Bae W.C., Masuda K., Chung C.B., Hwang D. Fine-grain segmentation of the intervertebral discs from MR spine images using deep convolutional neural networks: BSU-Net. *Applied Sciences*. 2018. V. 8 (9). P. 1656.  
<https://doi.org/10.3390/app8091656>
- Lustberg T., van Soest J., Jochems A., Deist T., van Wijk Y. Big Data in radiation therapy: challenges and opportunities. *The British journal of radiology*. 2017. V. 90 (1069). 20160689.  
<https://doi.org/10.1259/bjr.20160689>
- Mirikhharaji Z., Yan Y., Hamarneh G. Learning to segment skin lesions from noisy annotations. *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*. Cham. Springer. 2019. P. 207–215.  
[https://doi.org/10.1007/978-3-030-33391-1\\_24](https://doi.org/10.1007/978-3-030-33391-1_24)
- Nie D., Gao Y., Wang L., Shen D. Asdnet: Attention based semi-supervised deep networks for medical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham. Springer. 2018. P. 370–378.  
[https://doi.org/10.1007/978-3-030-00937-3\\_43](https://doi.org/10.1007/978-3-030-00937-3_43)
- Otsu N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*. 1979. V. 9 (1). P. 62–66.  
<https://doi.org/10.1109/tsmc.1979.4310076>
- Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Cham. Springer. 2015. P. 234–241.  
[https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Sahiner B., Pezeshk A., Hadjiiski L.M., Wang X., Drukker K. Deep learning in medical imaging and radiation therapy. *Medical physics*. 2019. V. 46 (1). P. e1–e36.  
<https://doi.org/10.1002/mp.13264>
- Sandström H., Jokura H., Chun C., Toma-Dasu I. Multi-institutional study of the variability in target delineation for six targets commonly treated with radiosurgery. *Acta Oncologica*. 2018. V. 57 (11). P. 1515–1520.  
<https://doi.org/10.1080/0284186x.2018.1473636>
- Tajbakhsh N., Jeyaseelan L., Li Q., Chiang J.N., Wu Z. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*. 2020. 101693.  
<https://doi.org/10.1016/j.media.2020.101693>
- Tanaka D., Ikami D., Yamasaki T., Aizawa K. Joint optimization framework for learning with noisy labels. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. P. 5552–5560.  
<https://doi.org/10.1109/CVPR.2018.00582>
- Vinod S.K., Jameson M.G., Min M., Holloway L.C. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. *Radiotherapy and Oncology*. 2016. V. 121 (2). P. 169–179.  
<https://doi.org/10.1016/j.radonc.2016.09.009>
- Wang C., Zhu X., Hong J.C., Zheng D. Artificial intelligence in radiotherapy treatment planning: present and future. *Technology in cancer research & treatment*. 2019. V. 18.  
<https://doi.org/10.1177/1533033819873922>
- Xue C., Dou Q., Shi X., Chen H., Heng P.A. Robust learning at noisy labeled medical images: applied to skin lesion classification. *IEEE 16th International Symposium on Biomedical Imaging*. IEEE. 2019. P. 1280–1283.  
<https://doi.org/10.1109/isbi.2019.8759203>
- Zeng Q., Karimi D., Pang E.H.T., Mohammed S., Schneider C. Liver Segmentation in Magnetic Resonance Imaging via Mean Shape Fitting with Fully Convolutional Neural Networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham. Springer. 2019. P. 246–254.  
[https://doi.org/10.1007/978-3-030-32245-8\\_28](https://doi.org/10.1007/978-3-030-32245-8_28)
- Zhang C., Bengio S., Hardt M., Recht B., Vinyals O. Understanding deep learning requires rethinking generalization. *5th International Conference on Learning Representations, Conference Track Proceedings*. Toulon. OpenReview.net. 2017a.
- Zhang Y., Yang L., Chen J., Fredericksen M., Hughes D.P. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham. Springer. 2017b. P. 408–416.  
[https://doi.org/10.1007/978-3-319-66179-7\\_47](https://doi.org/10.1007/978-3-319-66179-7_47)
- Zhu H., Shi J., Wu J. Pick-and-learn: Automatic quality evaluation for noisy-labeled image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham. Springer. 2019. P. 576–584.  
[https://doi.org/10.1007/978-3-030-32226-7\\_64](https://doi.org/10.1007/978-3-030-32226-7_64)
- Zhao F., Li M., Kong L., Zhang G., Yu J. Delineation of radiation therapy target volumes for patients with postoperative glioblastoma: A review. *Onco Targets Ther*. 2016. V. 9. P. 3197–3204.  
<https://doi.org/10.2147/ott.s104241>