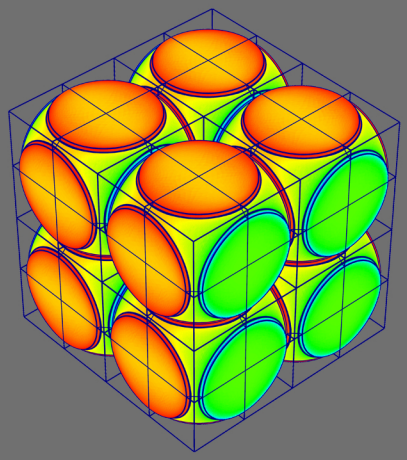
ISSN 2309-3684



Математическое моделирование и численные методы

Валишин А.А., Запривода А.В., Цухло С.С. Повышение эффективности обучения нейронных сетей для сегментации изображений. Математическое моделирование и численные методы, 2025, № 3, с. 103–116.

Источник: https://mmcm.bmstu.ru/articles/385/

Параметры загрузки:
IP: 216.73.216.62
23.11.2026 00:07:00

УДК 519.7 DOI: 10.18698/2309-3684-2025-3-103116

Повышение эффективности обучения нейронных сетей для сегментации изображений

© А.А. Валишин[1], А.В. Запривода[2], С.С. Цухло[1]

[1]МГТУ им. Н.Э. Баумана, Москва, 105005, Россия
[2]АО «ВПК «НПО машиностроения», Реутов, 143966, Россия

В данной статье критически анализируются современные подходы к повышению эффективности обучения нейронных сетей для сегментации изображений. Под эффективностью обучения понимаются два взаимосвязанных аспекта: вычислительная эффективность и точность сегментации обученной модели. Особое внимание уделено трём способам повышения эффективности обучения:1. применение методов аугментации. Аугментация – это процесс искусственного генерирования новых данных на основе существующих для обучения новых моделей. Этот метод позволяет увеличить размер и разнообразие набора данных, что важно для улучшения обобщающей способности модели. В данном случае аугментация включала в себя повороты изображения, наложение гауссовского шума, коррекция цветовой гаммы, 2. оптимизация архитектур нейронных сетей посредством интеграции эффективных энкодеров на базе EfficientNet, 3. применение методов активного обучения для выбора наиболее информативных обучающих примеров на основе вычисления энтропии выходных данных. Модели обучались с использованием оптимизатора Adam на задаче OpenEarthMap, где выборка составляла 20 % от исходного объёма, изображения уменьшались до разрешения 512×512 пикселей и дополнительно разбивались на 4 части размером 256×256 пикселей. Обучение проводилось на 9212 изображениях обучающей выборки и 1536 изображениях валидационной выборки в течение 100 циклов обучения. Результаты экспериментов показывают, что аугментация увеличивает точность сегментации модели UNet (IoU) с 36 % до 38,7 %, оптимизация архитектуры с использованием EfficientNet-b0 и b4 повышает IoU до 44,6 % и 45,3 % соответственно, а активное обучение, основанное на вычислении энтропии, демонстрирует потенциал выравнивания IoU по классам, хотя стабильность метрик остаётся проблематичной. Данная работа подчёркивает необходимость и перспективность комплексного подхода к оптимизации нейросетевых моделей для сегментации изображений и указывает направления для дальнейших исследований в области машинного обучения и повышения вычислительной эффективности.

Ключевые слова: *сверточные нейронные сети, компьютерное зрение, семантическая сегментация, машинное обучение, активное обучение*

Введение. Сегментация изображений представляет собой одну из важнейших задач компьютерного зрения, поскольку она позволяет разбивать изображение на смысловые участки, что существенно упрощает последующий анализ и принятие решений. Однако, несмотря на значительный прогресс, существующие методы сегментации сталкиваются с рядом фундаментальных проблем, влияющих на их точность и универсальность:

1) Неоднородность входных данных и влияние шума: Современные методы сегментации сталкиваются с большой вариативностью изображений: различия в освещении, наличии шума, размытости и артефактах негативно сказываются на качестве выделения объектов. Как отмечено в исследованиях [2], шум и вариативность условий съёмки затрудняют корректное разделение изображения на однородные области, что приводит к снижению точности сегментации.

2) Ограниченность и стоимость разметки данных: Для обучения моделей требуется большое количество качественно размеченных данных, где каждому пикселю присвоена корректная метка. Ручное аннотирование изображений – процесс трудоемкий, дорогой и подверженный субъективным ошибкам. Это ограничивает возможность масштабного обучения и негативно сказывается на обобщающей способности моделей [3,4].

3) Высокая вычислительная сложность современных архитектур: Мощные нейросетевые модели, демонстрирующие высокую точность сегментации, требуют значительных вычислительных ресурсов для обучения и использования модели. Это ограничивает их применение в системах реального времени и на устройствах с ограниченными ресурсами, что является существенным препятствием для внедрения в промышленные и мобильные приложения [5].

Кроме того, в работе [6] качественно продемонстрировано, что точность сегментации напрямую влияет на надёжность алгоритмов поиска изображений. Более чёткое выделение объектов позволяет значительно сократить количество ложных совпадений и повысить полноту поиска. Эти наблюдения подчёркивают важность разработки методов сегментации, способных обеспечивать высокую консистентность и точность разметки в разнообразных условиях съёмки.

Оптимизация нейросетевой архитектуры. Одной из важнейших задач при разработке систем сегментации изображений является не только улучшение точности предсказаний, но и оптимизация вычислительной эффективности модели. Архитектура U-Net [7] зарекомендовала себя как эффективное решение для сегментации, однако её прямое применение в задачах, требующих работы в реальном времени или на устройствах с ограниченными вычислительными ресурсами, может быть затруднено из-за большого числа параметров и вычислительных затрат. Для оптимизации архитектуры в данной работе предлагается замена стандартного энкодера U Net на более эффективные блоки на базе архитектуры EfficientNet [8]. EfficientNet, разработанный с использованием принципа масштабирования глубины, ширины и разрешения, обеспечивает высокий баланс между точностью и числом параметров. В рамках оптимизации были исследованы два варианта – модели с EfficientNet-b0 и EfficientNet-b4, которые применяются в качестве энкодера для сегментационных сетей.

Оптимизация нейросетевой архитектуры посредством интеграции EfficientNet в качестве нкодера имеет ряд преимуществ:

1) Сокращение числа параметров: Использование EfficientNet-b0 позволяет добиться значительного уменьшения числа параметров, что снижает требования к памяти и вычислительным ресурсам. Это особенно важно для применения моделей в реальном времени.

2) Увеличение скорости: более легкая архитектура обеспечивает более быструю обработку изображений без существенной потери точности, что критично для систем, где время реакции является ключевым параметром.

3) Повышение точности: несмотря на уменьшение числа параметров, оптимизированные модели демонстрируют улучшенную способность извлекать релевантные признаки, что приводит к повышению метрики IoU. Применение EfficientNet-b4, несмотря на сохранение сложности, позволяет достигать еще более высоких показателей сегментации благодаря более глубокому представлению данных.

4) Гибкость и масштабируемость: архитектура EfficientNet легко масштабируется в зависимости от требований приложения. Возможность выбора между более легкими и более сложными вариантами позволяет адаптировать модель под конкретные задачи, учитывая баланс между скоростью и точностью.

Методы обучения. При разработке нейросетевых моделей для сегментации изображений важна комплексная стратегия, которая включает несколько этапов: расширение обучающего набора с помощью аугментации, использование специализированной функции потерь для борьбы с дисбалансом классов, применение адаптивного оптимизатора и интеграция активного обучения для повышения эффективности обучения.

Аугментация данных – этап, позволяющий искусственно расширить объём обучающего датасета и повысить обобщающую способность модели. В данной работе применяется ряд строгих методов аугментации, среди которых основное внимание уделено фиксированным поворотам изображений.

1) Фиксированные повороты. Изображения поворачиваются на 0°, 90°, 180° или 270°, что позволяет модели увидеть объекты в разных ориентациях без дополнительной сложности случайного выбора углов. Преобразование координат для 90° поворота, например, определяется следующими формулами:

Для поворота изображения можно использовать матрицы преобразования. Пусть (x, y) — координаты исходного пикселя, тогда новые координаты (x', y') вычисляются по формулам (1) следующим образом для каждого угла (поворот системы координат производится по часовой стрелке):

$$\begin{aligned} 0°: x' = x, \quad y' = y, \\ 90°: x' = -y, \quad y' = x, \\ 180°: x' = -x, \quad y' = -y, \\ 270°: x' = y, \quad y' = -x. \end{aligned} \tag{1}$$

2)Добавление гауссовского шума. Для повышения устойчивости модели к изменениям условий съёмки, каждому пикселю изображения добавляется гауссовский шум. Если $I(x, y)$ — исходное значение пикселя, новое значение после добавления шума определяется как:

$$I_{noise}(x, y) = I(x, y) + \mathcal{N}(0, \sigma^2), \tag{2}$$

где $\mathcal{N}(0, \sigma^2)$ — гауссовское распределение с математическим ожиданием 0 и дисперсией σ^2. Такой подход помогает модели лучше обучаться, учитывая естественные шумы и помехи, присутствующие в реальных данных. При обучении значение σ^2 варьировалось от 0 до 0,04.

3) Цветовая коррекция в пространстве HSV [9] (hue saturation value – тон насыщенность яркость). Важным элементом аугментации является корректировка цветовых характеристик изображений. Для этого исходное изображение преобразуется из цветового пространства красного, зеленого, синего (RGB) в пространство HSV. Затем случайным образом изменяются значения оттенка, насыщенности или яркости, что позволяет модели обучаться на данных, полученных в различных условиях освещения. После этих изменений изображение возвращается в цветовое пространство RGB для дальнейшей обработки. Визуально пространство HSV можно представить как на рис. 1.

Пространство HSV представляет собой цилиндрическую систему координат, где координата H является полярным углом, S — полярный радиус, V — Z-координата. Переход между пространством RGB и HSV осуществляется по формулам (3):

$$H=\begin{cases} 60\dfrac{G-B}{R-B}+0, & \text{если } \max\left(R,G,B\right)=R,\ G\geq B \\ 60\dfrac{G-B}{R-G}+360, & \text{если } \max\left(R,G,B\right)=R,\ G<B \\ 60\dfrac{B-R}{G-\min\left(R,B\right)}+120, & \text{если } \max\left(R,G,B\right)=G \\ 60\dfrac{R-G}{B-\min\left(R,G\right)}+240, & \text{если } \max\left(R,G,B\right)=B \end{cases}$$

$$S=\begin{cases} 0, & \text{если } \max\left(R,G,B\right)=0 \\ 1-\dfrac{\min\left(R,G,B\right)}{\max\left(R,G,B\right)}, & \text{иначе} \end{cases} \qquad (3)$$

$$V=\max\left(R,G,B\right),$$

где

$$H\in\left[0,360\right],\ R,G,B,S,V\in\left[0,1\right].$$

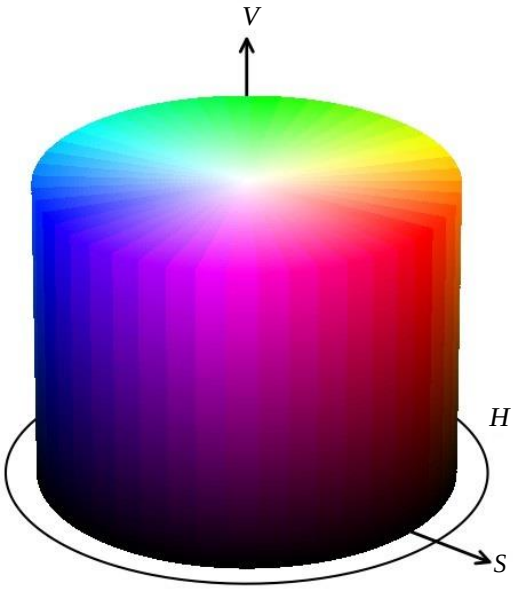


Рис. 1. Визуальное представление распределения цветов в пространстве HSV. H (hue, оттенок) — полярный угол, S (saturation, насыщенность) — полярный радиус, V (value, яркость) — Z-координата

Обратное преобразование проводится следующим образом: вычисляются значения $V,V_{min},V_{inc},V_{dec}$ по формулам (4). Затем эти зна-

чения используются для определения цвета в пространстве RGB согласно таблице 1.

$$V_{min} = (1 - S)V,$$
$$a = (V - V_{min})\frac{H \bmod 60}{60}, \quad (4)$$
$$V_{inc} = V_{min} + a; \quad V_{dec} = V - a.$$

Таблица 1

Таблица подстановок для определения цвета в пространстве RGB

$\left\lfloor \frac{H}{60} \right\rfloor \bmod 6$	**R**	**G**	**B**
0	V	V_{inc}	V_{min}
1	V_{dec}	V	V_{min}
2	V_{min}	V	V_{inc}
3	V_{min}	V_{dec}	V
4	V_{inc}	V_{min}	V
5	V	V_{min}	V_{dec}

В процессе обучения цвет изображений в пространстве HSV корректировался случайным образом на следующие величины:

$$\Delta H \in [-30; +30],\ \Delta S \in [-0{,}15; +0{,}15],\ \Delta V \in [-0{,}15; +0{,}15].$$

Функция потерь – это математическая функция, которая измеряет разницу между предсказанием нейронной сети и истинным ответом. Одной из проблем сегментации является дисбаланс классов: часто некоторые классы объектов занимают большую часть изображения, тогда как другие классы – меньшую. Для борьбы с этой проблемой используется функция фокусной потери (focal loss) [10], которая модифицирует стандартную кросс-энтропию за счёт введения модифицирующего множителя, уменьшающего вклад легко классифицируемых примеров.

Формула фокусной потери для одного примера выглядит следующим образом:

$$\mathcal{L}_{focal} = -\alpha (1 - \hat{p})^{\gamma} \log(\hat{p}), \quad (5)$$

где $\hat{p}$ — предсказанная вероятность для правильного класса, α — весовой коэффициент для балансировки классов, задается картой весов, вычисленной для каждого класса [7], γ — параметр фокусировки, использовалось $\gamma = 2$.

При обучении сегментационной модели функция потерь рассчитывается для каждого пикселя, а итоговая ошибка усредняется по всем пикселям. Таким образом, функция фокусной потери позволяет модели сконцентрироваться на трудных примерах, улучшая качество сегментации особенно для объектов, встречающихся реже.

Для обновления весов модели применялся метод градиентного спуска Adam [11] (Градиентный спуск с адаптивной оценкой момента), который адаптивно масштабирует шаг обновления для каждого параметра. Он использует экспоненциально взвешенное скользящее среднее градиентов и их квадратов. При обучении величины коэффициентов затухания использовались следующие: $\beta_1 = 0.9;\quad \beta_2 = 0.999$.

Активное обучение – это подход в машинном обучении, при котором сама модель выбирает, какие данные ей стоит изучить в первую очередь. Идея в том, что не все примеры из обучающей выборки одинаково полезны: некоторые – простые, а другие – сложные, граничные или неоднозначные. Активное обучение направлено на улучшение качества модели за счёт выбора наиболее информативных примеров для аннотации и последующего обучения. В контексте сегментации изображений используется критерий энтропии для оценки уверенности модели в её предсказаниях. Энторопия сегментированного изображения характеризует степень неопределенности в предсказаниях модели.

$$H(p) = -\sum_{j=1}^{C} p_j \log p_j, \tag{6}$$

где p_j — вероятность, предсказанная для класса сегментации j; C — количество классов сегментации. Примеры с высокой энтропией, то есть те, где модель не уверена в предсказании, формируются в выборку с целью проведения на ней цикла обучения модели. Таким образом, активное обучение позволяет эффективно использовать ресурсы и улучшить общую способность модели обрабатывать сложные и неоднозначные случаи, что в конечном итоге приводит к повышению точности сегментации.

Оценка точности сегментации проводилось при помощи метрики IoU (Intersection over Union) известной как коэффициент Жаккара, вычисляемой по формуле (7). Идея состоит в том, что находится отношение числа пересечений множества предсказанных моделью сегментов изображения и множества размеченных сегментов изображения к числу обединения этих множеств.

$$IoU = \frac{|P \cap T|}{|P \cup T|}, \tag{7}$$

где P — (prediction) предсказанная карта сегментов модели, T — (target) разметка сегментов изображения.

Обучение моделей проводилось на наборе данных OpenEarthMap [12]. Набор предназначен для обучения нейронных сетей для задачи сегментации спутниковых фотоснимков. Каждый снимок сопровождается сегментированным изображением или разметкой. Все объекты на изображениях разделены на 8 классов: земля, поле, развитое пространство, дороги, деревья, водоемы, С/Х земля, здания.

Для сокращения времени обучения и оптимизации использования вычислительных ресурсов изображения уменьшались до размера 512x512 пикселей и разбивались на 4 части размера 256×256. Из этого набора были сформированы: выборка из 9212 обучающих примеров и выборка из 1536 валидационных примеров. Также для проведения одного цикла обучения использовалось только 20 % данных из обучающей выборки, выбираемых равномерно либо методом активного обучения.

Результаты. Сначала оценивалось влияние методов аугментации данных на качество сегментации при использовании архитектуры UNet. Модель обучалась в течение 100 циклов с величиной шага оптимизации 5×10^{-5}. На рис. 2 приведены кривые динамики метрики IoU и значений фокусной потери в процессе обучения. Эксперимент показал, что применение аугментации обучающей выборки позволило повысить показатель IoU с 36 % до 38,7 %. При этом в обоих случаях наблюдалось устойчивое снижение значения фокусной потери, что свидетельствует о сходимости параметров модели к оптимуму и отсутствии признаков переобучения. Для всех прогонов использовались одинаковые значения начальных весов сети.

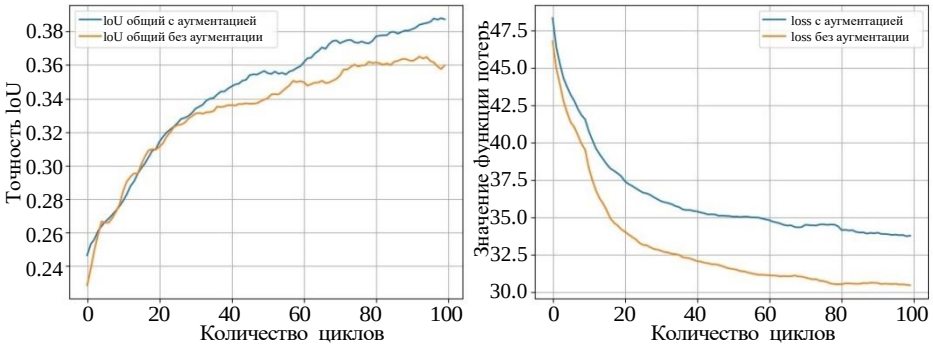


Рис. 2. Сравнение метрик при обучении модели UNet с аугментацией и без аугментации за 100 циклов обучения: IoU (слева), значение фокусной потери (справа). Использовалось скользящее среднее по последним 10 значениям

Далее было проведено сравнительное исследование трёх архитектур — EfficientUNet-b0, EfficientUNet-b4 и классической UNet — с применением того же метода аугментации, на протяжении 100 циклов и при шаге обучения 5×10^{-5}. На рис. 3 представлены кривые ди-

намики метрики IoU и значений фокусной потери в процессе обучения. Результаты показывают, что EfficientUNet-b0 достигает IoU 44,6 %, а EfficientUNet-b4 — 45,3 %, тогда как для UNet этот показатель составляет лишь 38,7 %. При этом фокусная потеря для моделей EfficientUNet демонстрирует более быстрое и глубокое снижение, достигая более низких значений по сравнению с UNet, что свидетельствует о лучшей сходимости и повышенной устойчивости к переобучению у архитектур на базе EfficientNet.

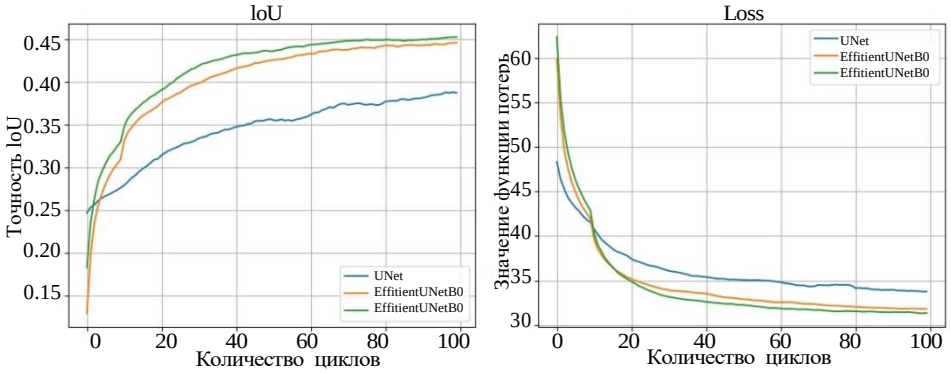


Рис. 3. Сравнение метрик при обучении моделей UNet, EfficientUNet-b0 и EfficientUNet-b4 за 100 циклов обучения: IoU (слева), значение фокусной потери (справа). Использовалось скользящее среднее по последним 10 значениям

Кроме повышения качества сегментации, архитектуры EfficientUNet показали и значительное преимущество по скорости обучения. При прочих равных условиях время полного обучения классической UNet составило примерно 5,5 ч, тогда как EfficientUNet-b0 потребовал около 2,5 ч, а EfficientUNet-b4 – около 5 ч. Таким образом, EfficientUNet-b0 обеспечивает более чем двукратное ускорение обучения по сравнению с UNet при одновременном росте метрики IoU, тогда как EfficientUNet-b4 сочетает близкие к UNet затраты времени с наивысшей точностью.

Затем с использованием модели EfficientUNet-b0 было исследовано применение стратегии активного обучения, при которой отбор обучающих примеров осуществлялся на основе максимальной средней энтропии выходных распределений модели. Обучение проводилось при тех же условиях (100 циклов, шаг обучения = 5×10^{-5}), на рисунке 4 представлены кривые динамики метрики IoU и функции фокусной потери. Итоговый средний IoU при активном отборе составил 36,5 % против 44,6 % при пассивном (случайном) выборе примеров; при этом кривые метрик, даже после обработки скользящим средним, демонстрируют сильную флуктуацию и отсутствие устойчивой монотонной сходимости, что указывает на нестабильность

процесса оптимизации. В то же время при анализе метрик IoU по отдельным классам обнаруживаются следующие закономерности:

1) Для наиболее труднообучаемых классов активный метод обеспечивает появление первых корректных предсказаний уже на начальных эпохах обучения (рис.5), что свидетельствует о частичном решении проблемы несбалансированности выборки за счёт концентрации на наиболее «сложных» примерах.

2) Межклассовая дисперсия итоговых значений точности при активном обучении ниже, чем при пассивном (таблица 2), что указывает на более равномерное качество сегментации по всем классам.

Таблица 2

Точность IoU по классам при обучении пассивным и активным методами.

Метод	Пассивный	Активный
Земля	14,55 %	9,35 %
Поле	35,46 %	28,50 %
Развитое пространство	33,92 %	28,69 %
Дороги	38,13 %	25,79 %
Деревья	46,45 %	41,73 %
Водоемы	24,26 %	20,52 %
С/Х земля	46,94 %	38,73 %
Здания	53,72 %	40,30 %
σ^2	143,29	106,6

Однако применение активного подхода оказалось значительно более ресурсоёмким: при том же объёме обучающих данных общее время обучения выросло до 30 ч, что ставит под вопрос практическую целесообразность метода.

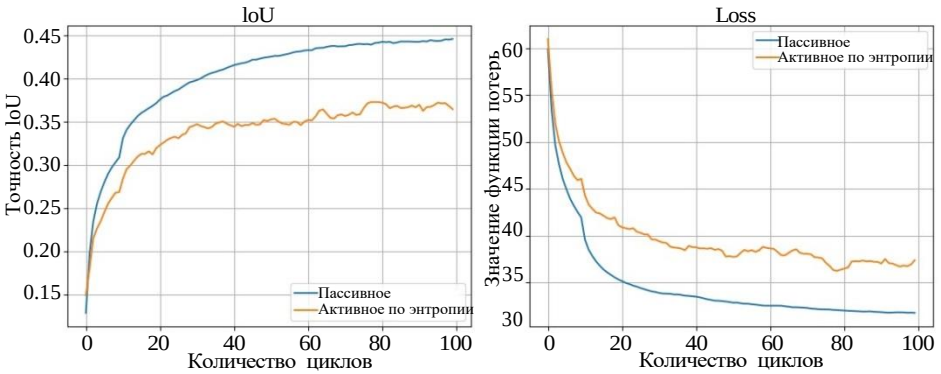


Рис. 4. Сравнение метрик при обучении модели EfficientUNet-b0 при пассивном и активном выборе обучающих примеров за 100 циклов обучения: IoU (слева), значение фокусной потери (справа). Использовалось скользящее среднее по последним 10 значениям

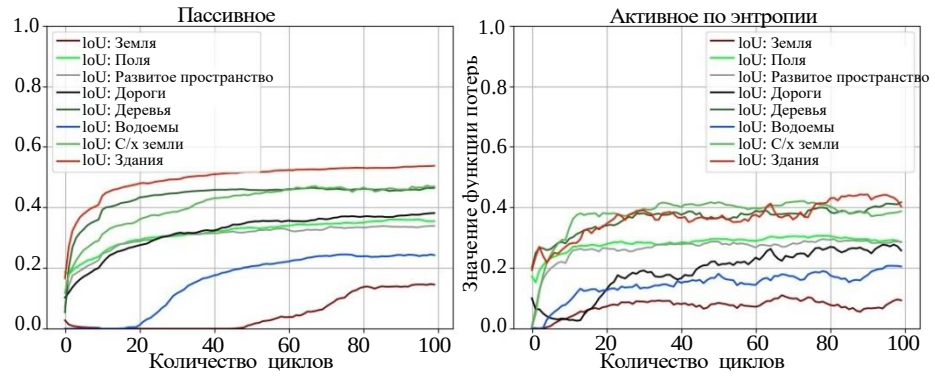


Рис. 5. Сравнение метрик IoU по каждому классу при обучении модели EfficientUNet-b0 при пассивном и активном выборе обучающих примеров за 100 циклов обучения: Пассивный метод (слева), Активный метод (справа). Использовалось скользящее среднее по последним 10 значениям

Заключение. В ходе исследования были сравнительно оценены три ключевых направления повышения качества семантической сегментации на архитектуре UNet. Применение техники аугментации позволило повысить средний IoU на 2,7 % при сохранении устойчивая монотонного снижения фокусной потери. Это свидетельствует о надёжной сходимости модели и отсутствии признаков переобучения. Архитектуры EfficientUNet-b0 и EfficientUNet-b4 показали существенное превосходство над классической UNet как по точности сегментации, так и по скорости обучения. Наиболее оптимальным с точки зрения соотношения «точность / ресурсоёмкость» является EfficientUNet-b0. Метод активного обучения с выбором наиболее «информативных» примеров по средней энтропии выходного распределения частично сгладила дисбаланс классов: появление первых корректных предсказаний для трудных классов на ранних эпохах; снижение межклассовой дисперсии IoU. Однако общая точность модели уменьшилась на 8,1 %, а время обучения выросло до 30 ч, что ограничивает практическую ценность метода без дополнительной оптимизации.

Таким образом, для приложений, предъявляющих высокие требования к качеству сегментации при ограниченных вычислительных ресурсах, рекомендуется сочетать аугментацию данных с архитектурой EfficientUNet-b0. Активное обучение целесообразно использовать в задачах, где важно выравнивание точности по всем классам, при условии внедрения механизмов стабилизации процесса обучения.

ЛИТЕРАТУРА

[1] Ulku I., Akagunduz E. A survey on deep learning-based architectures for semantic segmentation on 2D images. *Applied Artificial Intelligence*, 2022, vol. 36, no. 1. DOI:10.1080/08839514.2022.203292

[2] Валишин А.А., Запривода А.В., Клонов А.С. Математическое моделирование и сравнительный анализ численных методов решения задачи непрерывно-дискретной фильтрации случайных процессов в реальном времени. *Математическое моделирование и численные методы*, 2024, № 1, с. 93-109.

[3] Достовалова А.М. Моделирование локально-однородных радиолокационных изображений при использовании различных статистических критериев. *Математическое моделирование и численные методы*, 2021, № 4, с. 103-120.

[4] Кутыркин В.А., Чалей М.Б. Стохастические модели кодирования и распознавания структурно-статистических характеристик кодирующих последовательностей. *Математическое моделирование и численные методы*, 2017, № 3, с. 119-138

[5] Minaee S., Boykov Y., Porikli F., Plaza A., Kehtarnavaz N., Terzopoulos D. Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, vol. 44, no. 7, pp. 3523-3542.

[6] Валишин А.А., Запривода А.В., Цухло С.С. Моделирование и сравнительный анализ эффективности перцептивных хеш-функций для поиска сегментированных изображений. *Математическое моделирование и численные методы*, 2024, № 2, с. 46-67.

[7] Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *arXiv.org*. [Электронный ресурс]. URL: https://arxiv.org/abs/1505.04597 (дата обращения: 05.06.2024).

[8] Tan M., Le Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv.org*. [Электронный ресурс]. URL: https://arxiv.org/abs/1905.11946 (дата обращения: 23.05.2024).

[9] Yan Q., Feng Y., Zhang C., Wang P., Wu P., Dong W., Sun J., Zhang Y. You only need one color space: an efficient network for low-light image enhancement. *arXiv.org*. [Электронный ресурс]. URL: https://doi.org/10.48550/arXiv.2402.05809 (дата обращения: 12.05.2024).

[10] Lin T.-Yi, Goyal P., Girshick R., He K., Dollár P. Focal Loss for Dense Object Detection. *Conference: 2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. DOI:10.1109/ICCV.2017.324

[11] Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization. *arXiv.org*. [Электронный ресурс]. URL: https://doi.org/10.48550/arXiv.1412.6980 (дата обращения: 11.06.2024).

[12] Xia J., Yokoya N., Adriano B., Broni-Bediako C. OpenEarthMap: A Benchmark Dataset for Global High-Resolution Land Cover Mapping. *arXiv.org*. [Электронный ресурс]. URL: https://doi.org/10.48550/arXiv.2210.10732 (дата обращения: 24.05.2024).

Статья поступила в редакцию 20.06.2024

Ссылку на эту статью просим оформлять следующим образом:

Валишин А.А., Запривода А.В., Цухло С.С. Повышение эффективности обучения нейронных сетей для сегментации изображений. *Математическое моделирование и численные методы*, 2025, № 3, с. 103–116.

Валишин Анатолий Анатольевич — д-р физ.-мат. наук, профессор кафедры «Вычислительная математика и математическая физика» МГТУ им. Н.Э. Баумана. e-mail: enf@mail.ru

Запривода Алексей Васильевич — инженер АО «ВПК «НПО машиностроения».

Цухло Сергей Сергеевич — студент кафедры «Вычислительная математика и математическая физика» МГТУ им. Н.Э. Баумана.

Improving the training efficiency of neural networks for image segmentation

© A.A. Valishin[1], A.V. Zaprivoda[2], S.S. Tsukhlo[1]

[1]Bauman Moscow State Technical University, Moscow, 105005, Russia
[2]JSC «MIC «NPO Mashinostroyenia», 143966, Russia

This paper critically analyzes modern approaches to improving the efficiency of training neural networks for image segmentation. Training efficiency refers to two interrelated aspects: computational efficiency and segmentation accuracy of the trained model. Particular attention is paid to three ways to improve training efficiency: 1. using augmentation methods. Augmentation is the process of artificially generating new data based on existing models for training. This method allows increasing the size and diversity of the dataset, which is important for improving the generalization ability of the model. In this case, augmentation included image rotations, Gaussian noise imposition, color correction, 2. optimization of neural network architectures by integrating efficient encoders based on EfficientNet, 3. using active learning methods to select the most informative training examples based on calculating the entropy of the output data. The models were trained using the Adam optimizer on the OpenEarthMap task, where the sample was 20% of the original volume, the images were downsampled to 512x512 pixels and further split into four parts of 256x256 pixels. Training was performed on 9212 images of the training set and 1536 images of the validation set for 100 training cycles. The experimental results show that augmentation increases the segmentation accuracy of the UNet model (IoU) from 36 % to 38.7 %, architectural optimization using EfficientNet-b0 and b4 increases IoU to 44.6 % and 45.3 %, respectively, and active learning based on entropy calculation shows the potential to equalize IoU across classes, although the stability of the metrics remains problematic. This work highlights the need and potential of an integrated approach to optimizing neural network models for image segmentation and points to directions for further research in the field of machine learning and improving computational efficiency.

***Keywords**: convolutional neural networks, computer vision, semantic segmentation, machine learning, active learning*

REFERENCES

[1] Ulku I., Akagunduz E. A survey on deep learning-based architectures for semantic segmentation on 2D images. *Applied Artificial Intelligence*, 2022, vol. 36, no. 1. DOI:10.1080/08839514.2022.203292

[2] Valishin A.A., Zaprivoda A.V., Klonov A.S. Mathematical modeling and comparative analysis of numerical methods for solving the problem of continuous-discrete filtering of random processes in real time. *Mathematical Modeling and Computational Methods*, 2024, no. 1, pp. 93-109.

[3] Dostovalova A.M. Simulation of locally homogeneous radar images using different statistical criteria. *Mathematical Modeling and Computational Methods*, 2021, no. 4, pp. 103-120.

[4] Kutyrkin V.A., Chaley M.B. Stochastic coding models and the distribution of structural and statistical characteristics of coding sequences. *Mathematical Modeling and Computational Methods*, 2017, no. 3, pp. 119-138

[5] Minaee S., Boykov Y., Porikli F., Plaza A., Kehtarnavaz N., Terzopoulos D. Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, vol. 44, no. 7, pp. 3523-3542.

[6] Valishin A.A., Zaprivoda A.V., Tsukhlo S.S. Modeling and efficiency analysis of perceptual hash functions for segmented image search. *Mathematical Modeling and Computational Methods*, 2024, no. 2, pp. 46-67.

[7] Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *arXiv.org.* [Электронный ресурс]. URL: https://arxiv.org/abs/1505.04597 (дата обращения: 05.06.2024).

[8] Tan M., Le Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv.org.* [Электронный ресурс]. URL: https://arxiv.org/abs/1905.11946 (дата обращения: 23.05.2024).

[9] Yan Q., Feng Y., Zhang C., Wang P., Wu P., Dong W., Sun J., Zhang Y. You only need one color space: an efficient network for low-light image enhancement. *arXiv.org.* [Электронный ресурс]. URL: https://doi.org/10.48550/arXiv.2402.05809 (дата обращения: 12.05.2024).

[10] Lin T.-Yi, Goyal P., Girshick R., He K., Dollár P. Focal Loss for Dense Object Detection. *Conference: 2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. DOI:10.1109/ICCV.2017.324

[11] Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization. *arXiv.org*. [Электронный ресурс]. URL: https://doi.org/10.48550/arXiv.1412.6980 (дата обращения: 11.06.2024).

[12] Xia J., Yokoya N., Adriano B., Broni-Bediako C. OpenEarthMap: A Benchmark Dataset for Global High-Resolution Land Cover Mapping. *arXiv.org*. [Электронный ресурс]. URL: https://doi.org/10.48550/arXiv.2210.10732 (дата обращения: 24.05.2024).

Valishin A.A., Dr. Sc. (Phys. — Math.), Professor, Department of Computational Mathematics and Mathematical Physics, Bauman Moscow State Technical University. e-mail: enf@mail.ru

Zaprivoda A.V., Engineer JSC «MIC «NPO Mashinostroyenia».

Tsukhlo S.S., Student of Department of Computational Mathematics and Mathematical Physics, Bauman Moscow State Technical University.