

ДИ ЧЖАО, И ТАН, ПЕРЦЕВ Д.Ю., ГУРИНОВИЧ А.Б., КУПРИЯНОВА Д.В.

LANET ДЛЯ СЕГМЕНТАЦИИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

В данной работе представлена оригинальная модель LANet, предназначенная для улучшения результатов сегментации медицинских изображений и основанная на нейронной сети MobileViT. Разработанные и интегрированные блоки Efficient Fusion Attention и Adaptive Feature Fusion улучшают качество извлечения признаков и уменьшают избыточность данных. Эффективность представленных блоков подтверждена множеством экспериментов, включая оценку точности на различных наборах данных, на основе таких метрик, как Dice, Precision, Recall, mIoU, оценку производительности модели, а также исследование абляции.

Ключевые слова: LANet, Lightweight Attention Network, блок эффективного слияния внимания, блок адаптивного слияния признаков, механизм внимания

Введение

Сегментация медицинских изображений нашла свое применение в клинической практике, позволяя выполнить точный и автоматизированный анализ анатомических структур и патологий. Целью сегментации является разделение изображения на определенные регионы и выделение области интереса на основе таких атрибутов, как цвет, текстура и форма. Этот процесс является базовым для таких задач, как поиск очагов поражения, измерение конкретных тканей и 3D-реконструкция [1]. Исторически поиском патологий занимались высококвалифицированные специалисты (например, радиологи, патологоанатомы), выполняя анализ снимков вручную. Однако такой подход является трудоемким, подвержен дополнительным рискам, связанным, например, с усталостью, и не обладает гарантированной воспроизводимостью, что делает его непригодным для обработки больших объемов медицинских данных. Развитие классических алгоритмов сегментации существенно упростили работу специалистов, однако они требуют настройки необходимых параметров вручную (например, задание порога [2], выбор начальных точек для метода на основе выращивания регионов [3] и т. д.). С появлением и развитием специализированного программного обеспечения для обучения глубоких нейронных сетей, появлением относительно дешевых устройств для обучения, фокус исследований сместился в сторону применения методов глубокого обучения.

В сегментации медицинских изображений на сегодняшний день существует ряд нерешенных проблем:

– существующие модели сверточных нейронных сетей часто демонстрируют высокую вычислительную сложность, что препятствует их развертыванию на устройствах с ограниченными ресурсами,

в т. ч. в клинических приложениях для обработки в реальном времени [4];

– для достижения точных результатов сегментации важно эффективно выделять признаки, относящиеся к задаче, и при этом минимизировать их избыточность [5]. Кроме того, слияние низкоуровневых и высокоуровневых признаков для многих моделей нейронных сетей представляет собой серьезную проблему при работе с медицинскими изображениями, т. к. данный процесс должен быть адаптивным и контекстно-зависимым, чтобы учесть дополнительную информацию и тем самым повысить способность сети к точному восприятию и представлению входных признаков.

Для решения указанных проблем была разработана модель нейронной сети LANet (от англ. *Lightweight Attention Network*, облегченная сеть внимания), включающая следующие новые блоки:

– блок эффективного слияния внимания (от англ. *Efficient Fusion Attention*, *EFA*), задачей которого является нахождение контекстно-зависимых признаков и минимизация их избыточности;

– блок адаптивного слияния признаков (от англ. *Adaptive Feature Fusion*, *AFF*), который объединяет низкоуровневые и дискретизированные признаки, что улучшает их восприятие и представление сетью.

1. Модель нейронной сети LANet

При разработке модели нейронной сети LANet за основу была выбрана модель MobileViT [4], основанная на архитектуре трансформера, к которой дополнительно были добавлены блоки кодера и декодера. Кодер состоит из пяти блоков (на рис. 1 обозначены «E-Block»), каждый из которых соответству-

ет определенному разрешению карты признаков F_i ($i = 1, 2, \dots, 5$), размерность которой, в общем случае, равна $W/2^i \times H/2^i$. Несмотря на то, что эксперименты с MobileViT, проведенные другими исследователями, показали высокую эффективность данной модели, оказалось, что она может формировать избыточную информацию [6]. Поэтому каждый E-Block обрабатывает полученные ранее признаки с помощью EFA для выделения значимых на соответствующем уровне. Затем очищенные низкоуровневые карты объединяются с вышестоящим уровнем, результат объединения передается блоку декодирования AFF, целью которого является сохранить отличительные признаки, полученные в результате объединения информации от E-Block на разных уровнях, и расширить возможности их представления. Многие проанализированные методы ограничиваются только выходом самого верхнего слоя декодера в качестве результата предсказания, не обращая внимания на потерю пространственной информации в результате повышающей дискретизации блока декодирования. Чтобы решить эту проблему, выполняется преобразование масштаба с последующим объединением результатов AFF, полученные при разных разрешениях. В частности, чтобы учесть вариации формы и размера на различных уровнях при декодировании, выход каждого блока AFF подвергается глобальному усреднению. Кроме того, две различные нелинейные функции активации, т. е. ReLU и Sigmoid, применяются в связке со сверточными слоями 1×1 для оценки корреляции на разных уровнях и созда-

ния весов при обработке информации в цветовых каналах. Полученные веса затем умножаются на выходной сигнал, чтобы усилить значимость признаков, которые вносят существенный вклад в предсказание результата. Наконец, результаты, полученные в результате объединения четырех блоков декодирования, проходят через непрерывный сверточный слой 1×1 , слой повышающей дискретизации и еще один сверточный слой 1×1 для получения конечного результата предсказания.

Полная структура разработанной сети представлена на рис. 1 (базовая модель MobileViT на рисунке обозначена как «Backbone»). При этом приняты следующие обозначения:

– «Conv» обозначает сверточный слой с последующей пакетной нормализацией и функцией активации ReLU;

– синяя стрелка обозначает комбинацию сверточного слоя 1×1 и слоя повышающей дискретизации;

– 'U' и '+' обозначают слои повышающей дискретизации и поэлементного сложения, соответственно.

Блок эффективного слияния внимания (EFA) – это блок внимания, специально разработанный для повышения способности модели к использованию канальной и пространственной информации. На рис. 2 показана его структура, включающая 2 ветви:

– ветвь внимания к пространственной информации;

– ветвь внимания к канальной информации.

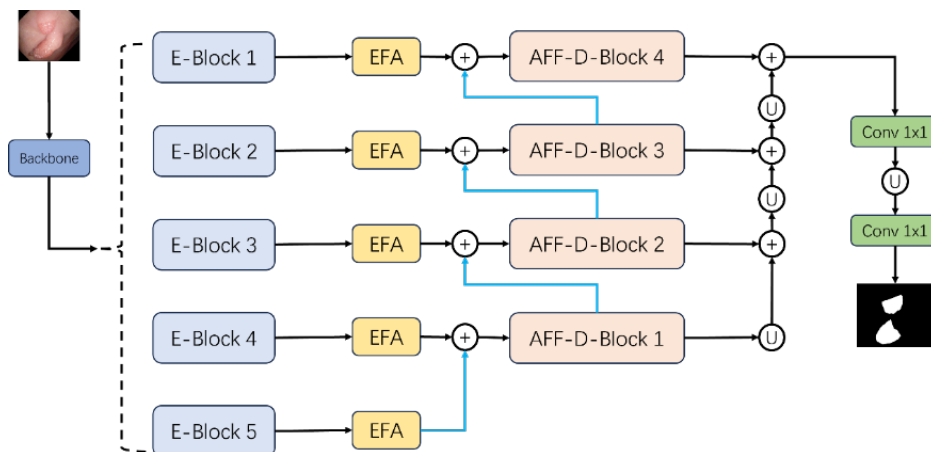


Рисунок 1. Модель нейронной сети LANet

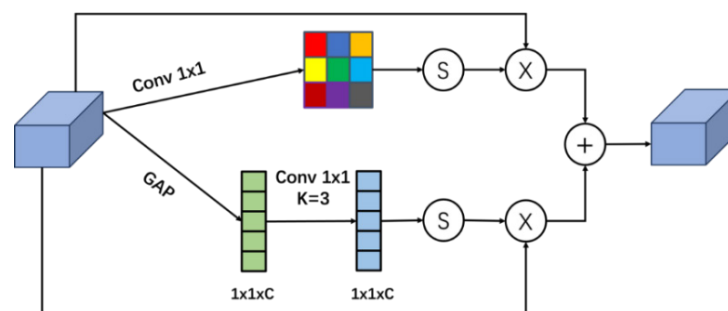


Рисунок 2. Структура блока EFA

Ветвь внимания к пространственной информации направлена на улучшение способности модели фокусироваться на пространственном расположении объектов в пределах входного изображения, применяя свертку 1×1 . Затем результаты свертки преобразуются в веса внимания в диапазоне от 0 до 1 с помощью сигмоидальной функции (на рис. 2 обозначена буквой S). В конечном итоге эти веса поэлементно умножаются на входные данные, чтобы подчеркнуть пространственные признаки, имеющие отношение к задаче, и ослабить остальные.

Ветвь внимания к каналной информации на основе входных данных формирует характеристики каналов. Она фиксирует их значимость с помощью последовательного выполнения таких операций, как глобальное усреднение входных данных (операция GAP) для получения общей информации о признаках, две операции свертки 1×1 для получения весов для каждого канала в отдельности. Подобно ветви внимания к пространственной информации, эти веса также сводятся к диапазону от 0 до 1 с помощью сигмоидальной функции и поэлементно умножаются на входные данные для выделения признаков в каналах, имеющих отношение к задаче, и подавления остальных.

Наконец, блок EFA объединяет выходы ветвей, складывая их поэлементно.

Такое слияние позволяет модели всесторонне использовать каналную и пространственную информацию, тем самым улучшая понимание и использование моделью входных признаков. Математически данный блок можно описать следующим образом:

$$\begin{aligned} O_c &= \text{Input} \cdot (\partial(\text{Conv}_{1 \times 1}(\text{GAP}(\text{Input})))) \\ O_s &= \text{Input} \cdot (\partial(\text{Conv}_{1 \times 1}(\text{Input}))) \\ O &= O_c + O_s \end{aligned} \quad (1)$$

где ∂ обозначает сигмоидальную функцию активации, GAP и $\text{Conv}_{1 \times 1}$ – слой глобального усреднения входных данных и сверточный слой 1×1 соответственно, Input – входное изображение.

Блок декодирования с адаптивным слиянием признаков (AFF). В разработанной модели кодирование играет важную роль в извлечении признаков из исходных данных, а декодирование отвечает за восстановление этих признаков к исходному разрешению изображения для формирования окончательной маски. Однако традиционная структура кодер-декодер может иметь определенные ограничения в контексте объединения признаков. Например, многие варианты модели U-Net [7-10] были направлены на совершенствование подходов к извлечению качественных признаков, но не рассматривали способы их объединения.

Блок AFF получает входную информацию, состоящую из очищенных низкоуровневых признаков кодера и признаков декодера. Эти два вектора признаков объединяются для формирования

слитого представления, которое затем разделяется на две параллельные ветви. Первая ветвь состоит из двух сверточных слоев 3×3 , каждый из которых проходит пакетную нормализацию и использует функцию активации ReLU. Эта ветвь предназначена для извлечения высокоуровневых признаков с помощью операции свертки, что позволяет модели улавливать более сложные и абстрактные объекты. Во второй ветви сначала используется сверточный слой 1×1 и слой пакетной нормализации для уменьшения размерности признаков и вычислительной сложности. Затем реализуется механизм внимания путем введения слоя Squeeze & Excitation. Этот слой обеспечивает адаптивное взвешивание признаков, позволяя модели выделять важные признаки и усиливать их. Наконец, выходы двух ветвей складываются и отображаются через сигмоидальную функцию для формирования конечного результата. Такой подход к объединению внимания и признаков демонстрирует повышенную гибкость и устойчивость.

Структура блока декодирования показана на рисунке 3.

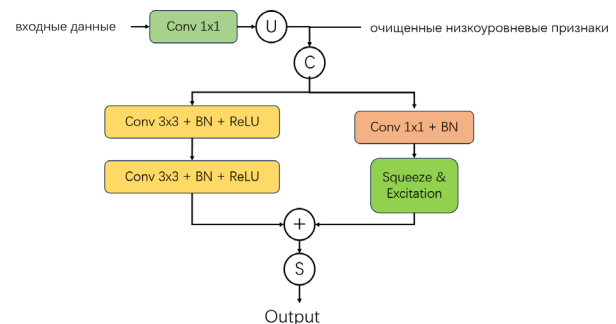


Рисунок 3. Структура блока AFF

Функция потерь. Бинарная кросс-энтропия – метод, широко используемый для оценки ошибки прогнозирования в различных задачах сегментации, однако она имеет ряд недостатков. Во-первых, метод не учитывает глобальную структуру изображения при оценке потерь для отдельных пикселей. Во-вторых, медицинские изображения в тестовых наборах несбалансированные и фоновых пикселей больше, чем граничных. Такой дисбаланс может привести к смещению в сторону предсказания категорий фона. Для решения указанных недостатков в LAnet предлагается использовать функцию потерь Lovasz-Softmax [11]:

$$\text{Loss} = \frac{1}{T} \sum_{t \in T} \overline{\Delta f}_t(m(t)) \text{ и } m_i(t) = \begin{cases} 1 - p_i(t), & t = y_i(t) \\ p_i(t), & \text{иначе} \end{cases} \quad (2)$$

где T и $(\overline{\Delta f}_t)$ обозначают число классов и расширение Ловаса для коэффициента Жаккара соответственно, а $y_i(t) \in \{-1, 1\}$ и $p_i(t) \in [0, 1]$ – истинная метка пикселя i и вероятность принадлежности пикселя i к классу t соответственно.

Результаты тестирования.

Оценка качества предложенной модели

Тестовые наборы данных. Для тестирования предложенной модели LANet использовались четыре открытых набора данных медицинских изображений:

– Kvasir-SEG [12]: открытый набор данных, состоящий из изображений полипов желудочно-кишечного тракта и соответствующих масок для сегментации. Аннотации были тщательно проработаны опытными гастроэнтерологами. Набор данных включает 1000 изображений полипов с разрешением от 332×487 до 1920×1072 пикселей;

– CVC-ClinicDB [13]: набор данных, содержащий 612 RGB-изображений с разрешени-

ем 384×288 пикселей и включающий различные структуры и поражения (например, нормальная ткань толстой кишки, полипы и язвы). В наборе данных учитываются различные пациенты и настройки объектива для обеспечения разнообразия и реалистичности;

– CVC-ColonDB [14]: специально созданный набор данных для анализа изображений колоноскопии и обнаружения полипов, состоящий из 380 изображений с разрешением 500×570 пикселей;

– 2018 Data Science Bowl [15]: набор данных, состоящий из 670 изображений с разрешением 256×256 пикселей, в котором представлены изображения биологических клеток (например, раковые, нормальные и другие).

Таблица 1. Параметры тестовых наборов, используемые для работы с LANet

Тестовый набор	Число изображений	Разрешение, пикселей	Результат деления тестовых наборов, изображений		
			тренировочный	проверочный	тестовый
Kvasir-SEG	1000	256 × 256	800	100	100
CVC-ClinicDB	612		490	61	61
CVC-ColonDB	380		300	40	40
2018 Data Science Bowl	670		530	67	67

При предварительной подготовке данных изображения были приведены к единому разрешению, равному 256×256 пикселей (для этого использовался метод линейной интерполяции с разными коэффициентами масштабирования по осям X и Y). Весь набор данных разделен на множества для обучения, проверки и тестирования в соотношении 8:1:1. В таблице 1 представлены результаты предварительной подготовки данных. Затем полученные карты признаков возвращались к исходному размеру для итоговой оценки результата.

Метрики оценки результатов. Для оценки эффективности работы LANet использовался ряд стандартных оценочных показателей: Precision (Prec.), Recall, Dice и mean Intersection over Union (mIoU):

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 Recall &= \frac{TP}{TP + FN} \\
 Dice &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \\
 mIoU &= \frac{1}{2} \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FN + FP} \right)
 \end{aligned} \quad (3)$$

где TP , FP , FN – число истинно положительных, ложно положительных и ложно отрицательных результатов классификации.

Особенности реализации. Модель нейронной сети реализована с применением фреймвор-

ка PyTorch и обучена с применением видеокарты NVIDIA GeForce RTX 4060 Ti с 16 ГБ памяти. В качестве опорной сети использовалась модель MobileVit, которая была предварительно обучена с помощью тестового набора ImageNet [16].

Для обучения модели LANet использовались следующие параметры:

- модель оптимизации: алгоритм Adam с начальной скоростью обучения $1e-4$;
- размер пакета (от англ. batch): 16;
- число эпох обучения: 300.

Чтобы решить проблему небольшого размера тестовых наборов, для их пополнения использовались следующие подходы: случайная обрезка изображений, повороты на $[90^\circ, 180^\circ, 270^\circ]$. Следует отметить, что увеличение данных не применялось к тестовому набору.

Оценка полученных результатов. В таблице 2 представлены основные результаты по выбранным метрикам в сравнении с другими современными моделями (U-Net [17], Attention U-Net [7], ResUNet++ [8], U-Net++ [9], DoubleU-Net [10], FANet [18], FCN8 [19], PraNet [20]). Для сравнения с разработанной моделью LANet для тестовых наборов Kvasir-SEG, CVC-ClinicDB и Data Science Bowl 2018 использовались результаты, представленные в работе [18]. Для тестового набора CVC-ColonDB использовались официальные реализации с сайта GitHub. Лучшие значения выделены жирным.

Таблица 2. Результаты работы с тестовым набором Kvasir-SEG

Тестовый набор	Модель нейронной сети	Dice	mIoU	Recall	Prec.
Kvasir-SEG	U-Net	0.597	0.471	0.617	0.672
	Attention U-Net	0.794	0.696	0.838	0.829
	ResUNet++	0.714	0.612	0.742	0.784
	U-Net++	0.800	0.700	0.872	0.799
	DoubleU-Net	0.813	0.733	0.840	0.861
	FANet	0.880	0.815	0.906	0.901
	FCN8	0.831	0.737	0.835	0.882
	Разработанная модель	0.911	0.851	0.903	0.949
CVC-ClinicDB	U-Net	0.823	0.755	-	-
	ResUNet++	0.796	0.796	0.702	0.879
	FANet	0.936	0.894	0.924	0.940
	PraNet	0.899	0.849	-	-
	Разработанная модель	0.944	0.896	0.926	0.966
CVC-ColonDB	U-Net	0.504	0.436	0.533	0.626
	U-Net++	0.482	0.408	0.525	0.607
	PraNet	0.712	0.640	0.749	0.852
	Разработанная модель	0.771	0.712	0.758	0.894
Data Science Bowl 2018	U-Net	0.757	0.910	-	-
	Attention U-Net	0.918	0.910	0.918	0.924
	U-Net++	0.912	0.848	0.920	0.911
	DoubleU-Net	0.768	0.841	0.641	0.960
	FANet	0.918	0.857	0.922	0.919
	Разработанная модель	0.930	0.871	0.918	0.946

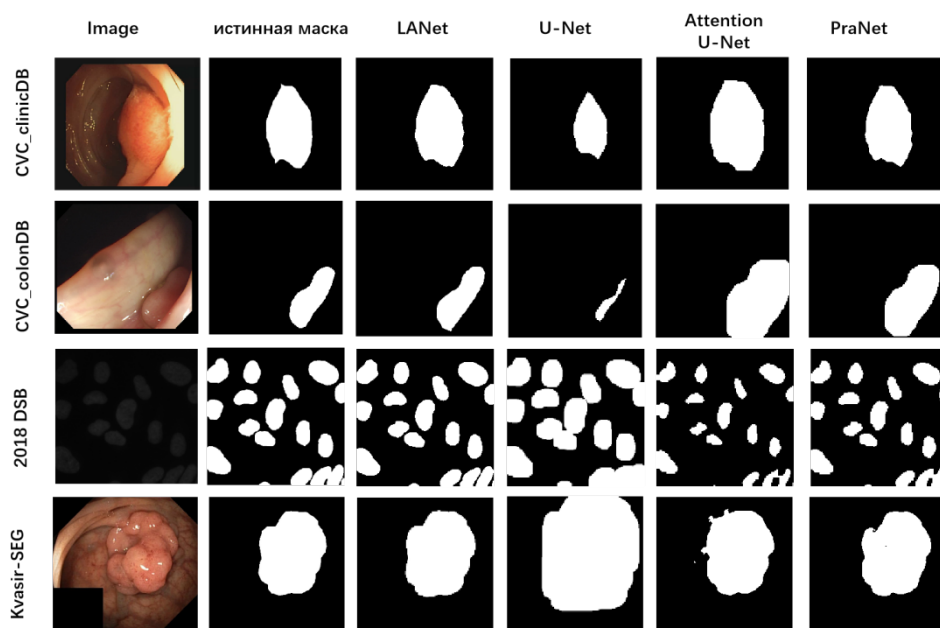


Рисунок 4. Визуализация результатов сегментации на примере 4 моделей нейронных сетей, включая разработанную

На рис. 4 представлены результаты сегментации, полученные при сравнении модели LANet с тремя другими методами. При этом видно, что представленная модель достигает близких к истине результатов, превосходя другие сравниваемые методы в работе в различных сложных ситуациях:

– небольшой размер области интереса в тесте CVC-ColonDB;

– несколько областей интереса небольшого размера с незначительными цветовыми изменениями между объектами и фоном в тесте Data Science Bowl 2018 (на рисунке 4 обозначен как «2018 DSB»);

– большой размер полипов и сложные нечеткие границы, связанные с особенностями освещения и теней.

Таблица 3. Результаты работы с тестовым набором Kvasir-SEG

Параметр	Unet	Unet++	PraNet	ResU-Net	MSNet	Attention U-Net	LANet
Число параметров, млн.	34.52	36.63	30.50	18.22	27.69	34.88	20.92
Производительность модели, GFLOPs	123.87	262.16	13.15	55.36	17.00	266.54	46.53
Скорость распознавания, fps	123.11	82.51	25.05	23.50	31.08	55.93	89.29

В таблице 3 представлена оценка сложности моделей нейронных сетей, на основе которой можно судить о сложности представленного решения в сравнении с альтернативными. В качестве основных параметров были выбраны следующие: размер модели в миллионах параметров, производительность модели, измеренная в GFLOPs, а также время распознавания в кадрах в секунду (fps).

Как видно из таблицы 3, предложенная модель занимает второе место по количеству параметров и по скорости распознавания по сравнению с другими методами. Сопоставление полученных результатов с результатами из таблицы 2 показывает, что модель LANet обеспечивает высокую скорость распознавания и производительность, сохраняя при этом высокий уровень точности.

Оценка t -критерия Стьюдента (на основе метрики Dice). Для набора данных Kvasir-SEG результаты оценки t -критерия показывают, что все модели, кроме FANet, имеют p -значения значительно ниже 0.05 (таблица 4), что говорит о том, что разница в полученной оценке метрики Dice по сравнению с моделью LANet статистически значима, т.е. наблюдаемые улучшения вряд ли объясняются случайностью. Однако сравнение с FANet показало p -значение, равное 0.41, что говорит об отсутствии статистически значимых различий. Такой результат связан с устойчивостью FANet к специфическим характеристикам изображений, присутствующим в тесте Kvasir-SEG, что может сделать его конкурентоспособным по отношению к предложенной модели при определенных условиях.

Таблица 4. Сравнение результатов проверки значимости моделей с помощью t -критерия на наборе данных Kvasir-SEG

Модель нейронной сети	U-Net	ResUNet++	FCN	DoubleU-Net	U-Net++	Attention U-Net	FANet
p -значение	2.453e-11	1.016e-05	0.047	0.024	0.015	0.004	0.411

Аналогично в тесте с базой CVC-ClinicDB результаты оценки t -критерия (таблица 5) показали статистически значимые p -значения для U-Net, ResUNet++, SFA и PraNet. При этом несмотря на то, что DoubleU-Net достигла значения Dice, равное 0.924, по сравнению с 0.944 в модели LANet, p -значение составило 0.910, что указывает на отсутствие статистически значимой разницы. Это говорит о том, что наблюдаемое различие может быть вызвано случайной вариацией, а не истинным различием в полу-

ченных картах признаков. Тем не менее, небольшое преимущество представленной модели в коэффициенте Dice может быть клинически значимым в некоторых приложениях, где даже незначительные улучшения имеют ценность.

В тесте с базой CVC-ColonDB результаты оценки t -критерия (таблица 5) демонстрируют статистическую значимость для всех сравниваемых моделей, что указывает на то, что полученные результаты не являются случайными.

В наборе данных Data Science Bowl 2018 представленная модель также демонстрирует лучшие коэффициенты Dice по сравнению с большинством методов. Однако сравнение с U-Net++, Attention U-Net и FANet показало p -значения 0.46, 0.89 и 0.79 соответственно, что не указывает на статистически значимые различия. Это говорит о том, что при работе с изображениями, на которых много объектов, предложенная модель суще-

ственно не превосходит эти методы. Отсутствие значимых различий можно объяснить сложностью и изменчивостью изображений в указанном наборе данных, что может уравнивать шансы U-Net++, Attention U-Net и FANet с предложенным подходом, подчеркивая необходимость дальнейшей оптимизации или применения специализированных подходов для достижения явного превосходства.

Таблица 5. Сравнение результатов проверки значимости моделей с помощью t -критерия на наборах данных CVC-ClinicDB, CVC-ColonDB и Data Science Bowl 2018

CVC-ClinicDB		CVC-ColonDB		Data Science Bowl 2018	
Модель нейронной сети	p -значение	Модель нейронной сети	p -значение	Модель нейронной сети	p -значение
U-Net	0.007	U-Net	1.67e-12	U-Net	0.001
ResUNet++	0.003	U-Net++	2.407e-10	DoubleU-Net	0.007
SFA	2.62e-07	C2FNet	0.049	U-Net++	0.463
PraNet	0.025	MSNet	0.037	Attention U-Net	0.887
Double U-Net	0.910	PraNet	0.018	FANet	0.786

Исследование абляции. Чтобы проверить значимость ключевых компонентов в модели LANet, был проведен ряд экспериментов по абляции на двух наборах данных (Kvasir-SEG и CVC-ClinicDB).

Результаты тестов приведены в таблице 6. Визуальное сравнение результатов сегментации представлено на рис. 5.

Удаление блока EFA не сильно сказывается на результатах измерения указанных метрик и не превышает 3,4 % в метрике Precision для тестового набора Kvasir-SEG и 3,3 % в метрике Dice для теста CVC-ClinicDB. Однако, анализируя рисунок 5, можно увидеть, что модель без EFA некорректно

обрабатывает края интересующих объектов, добавляя лишние детали.

Чтобы проанализировать вклад AFF, вместо данного блока были внедрены два сверточных слоя 3×3 с пакетной нормализацией и функцией активации ReLU в качестве блока декодирования. Как видно из таблицы 6, отсутствие блока AFF приводит к заметному снижению качества сегментации, особенно на наборе данных CVC-ClinicDB. В частности, метрики Dice, mIoU, Recall и Precision демонстрируют снижение на 3,7 %, 3,7 %, 3,3 % и 4,0 % соответственно. Кроме того, рис. 5 наглядно демонстрирует значительное расхождение между результатом работы модели без блока AFF и истинным значением.

Таблица 6. Оценка значимости основным компонент модели

Тестовый набор	Метод	Dice	mIoU	Recall	Prec.
Kvasir-SEG	без EFA	0.889	0.830	0.897	0.915
	без AFF	0.901	0.841	0.901	0.923
	LANet	0.911	0.851	0.903	0.949
CVC-ClinicDB	без EFA	0.911	0.865	0.903	0.958
	без AFF	0.907	0.859	0.893	0.926
	LANet	0.944	0.896	0.926	0.966

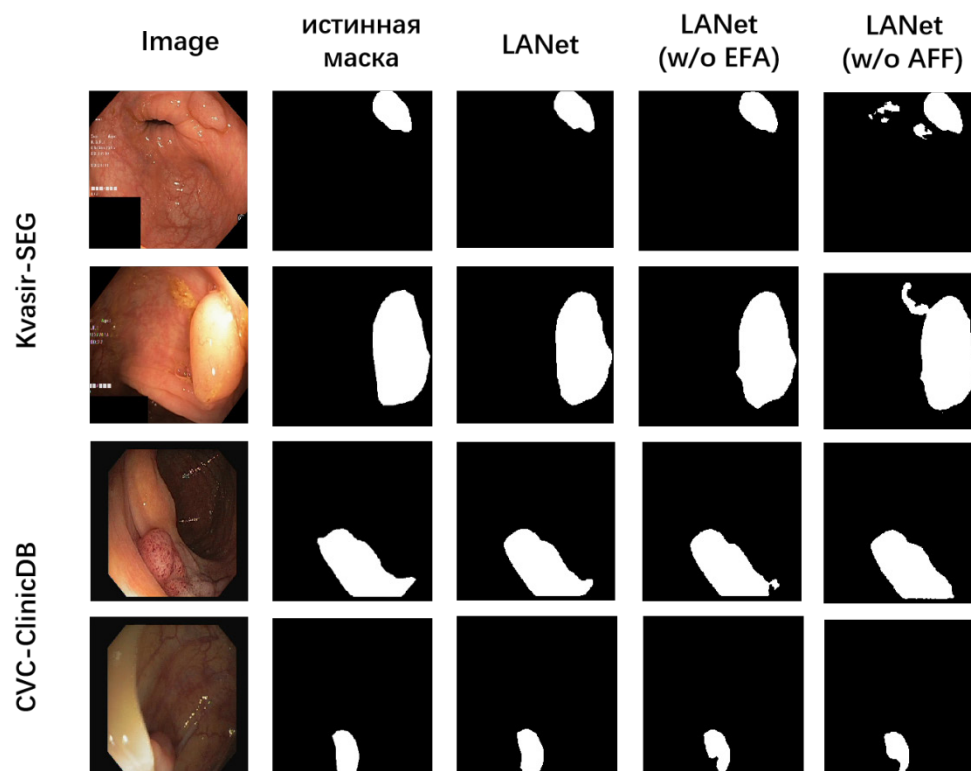


Рисунок 5. Визуальное сравнение результатов сегментации

Заключение

В данной работе представлена разработанная модель нейронной сети LANet для сегментации медицинских изображений. В качестве базовой модели используется MobileViT для эффективного извлечения информации о признаках из входных изображений. Предложены два новых компонента, составляющие основу модели LANet, – блок эффективного слияния внимания и блок декодирования с адаптивным слиянием признаков. Блок EFA улучшает возможности модели по извлечению признаков, захватывая релевантную задачу информацию и минимизируя избыточность как на канальном, так и пространственном уровнях. С другой стороны, блок декодирования AFF эффективно объединяет «очищенные» низкоуровневые признаки с признаками, полученными от декодера.

Для оценки эффективности сегментации LANet были проведены обширные исследования на четырех публичных наборах данных: Kvasir-SEG, CVC-ClinicDB, CVC-ColonDB и Data Science Bowl 2018. Результаты экспериментов показали, что предложенная модель превосходит другие современные методы. Кроме того, для исследования значимости ключевых компонентов были проведены эксперименты по абляции, результаты которых подтвердили эффективность предложенных блоков.

Анализ данных с использованием *t*-критерия Стьюдента показал, что предложенная модель демонстрирует статистически значимое улучшение

по сравнению с большинством моделей в различных наборах данных. Для тех параметров, где значимость не была достигнута, различия в метриках Dice были небольшими, однако в пользу предложенной модели.

Предложенные улучшения имеют важное значение в медицинской сфере, поскольку даже небольшое повышение точности сегментации важно в клинической практике для принятия точных решений. Это подчеркивает практическую ценность и эффективность модели LANet, благодаря ее способности обеспечивать качественные результаты сегментации в различных клинических сценариях.

В заключение следует отметить, что архитектура LANet с интегрированными блоками EFA и AFF представляет собой перспективное решение для задач сегментации медицинских изображений. Дальнейшие исследования могут быть направлены на совершенствование модели и изучение ее применимости в области трехмерных медицинских изображений.

По решению авторов, официальная реализация представленной модели была опубликована на сайте GitHub (<https://github.com/tyjcbzd/LANet>) и доступна для бесплатного некоммерческого использования.

Благодарности

Данная работа выполнена при поддержке Китайского стипендиального совета (CSC).

REFERENCES

1. **Taghanaki, S.A.** Deep semantic segmentation of natural and medical images: a review / S.A.Taghanaki, [etc.] // Artificial Intelligence Review. – 2021. – Vol. 54 – P. 137–178. – DOI: 10.1007/s10462-020-09854-1
2. **Tong, L.** A clustering-aided approach for diagnosis prediction: A case study of elderly fall / L. Tong, [etc.] // 2022 IEEE 46th Annual Computers, Software, and Applications Conference. – 2022. – P. 337–342. – DOI: 10.1109/COMPSAC54236.2022.00054
3. **Yu, H.** A scalable region-based level set method using adaptive bilateral filter for noisy image segmentation / H. Yu, F. He, Y. Pan // Multimedia Tools and Applications. – 2020. – Vol. 79. – P.5743-5765. – DOI: 10.1007/s11042-019-08493-1
4. **Mehta, S.** MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer / S. Mehta, M. Rastegari // arXiv preprint. – 2021. – arXiv: 2110.02178.
5. **Li, J.** SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy / J. Li, Y. Wen, L. He // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2023. – P. 6153-6162. – DOI: 10.1109/CVPR52729.2023.00596
6. **Qiu, J.** SlimConv: Reducing Channel Redundancy in Convolutional Neural Networks by Features Recombining / J. Qiu, [etc.] // IEEE Transactions on Image Processing. – 2021. – Vol. 30. – P. 6434-6445. – DOI: 10.1109/TIP.2021.3093795
7. **Oktay, O.** Attention U-Net: Learning Where to Look for the Pancreas / O. Oktay [etc.] // arXiv preprint. – 2018. – arXiv:1804.03999.
8. **Jha, D.** ResUNet++: An Advanced Architecture for Medical Image Segmentation / D. Jha, [etc.] // 2019 IEEE International Symposium on Multimedia (ISM). – 2019. – P. 225-235. – DOI: 10.1109/ISM46123.2019.00049
9. **Zhou, Z.** UNet++: A Nested U-Net Architecture for Medical Image Segmentation / Z. Zhou, [etc.] // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. – 2018. – P. 3-11. – DOI: 10.1007/978-3-030-00889-5_1
10. **Jha, D.** DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation / D. Jha [etc.] // 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). – 2020. – P. 558-564. – DOI: 10.1109/CBMS49503.2020.00111
11. **Berman, M.** The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks / M. Berman, T.A. Rannen, M.B. Blaschko // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2018. – P. 4413-4421. – DOI: 10.1109/CVPR.2018.00464
12. **Jham, D.** Kvasir-SEG: A Segmented Polyp Dataset / D. Jham, [etc.] // International Conference on Multimedia Modeling. – 2019. – P. 451-462. – DOI: 10.1007/978-3-030-37734-2_37
13. **Bernal, J.** WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians / J. Bernal, [etc.] // Computerized Medical Imaging and Graphics. – 2015. – Vol. 43. – P. 99-111. – DOI: 10.1016/j.compmedimag.2015.02.007
14. **Tajbakhsh, N.** Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information / N. Tajbakhsh, S.R. Gurudu, J. Liang // IEEE Transactions on Medical Imaging. – 2016. – Vol. 35, Issue 2. – P. 630-644. – DOI: 10.1109/TMI.2015.2487997.
15. **Caicedo, J. C.** Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl / J.C. Caicedo, [etc.] // Nature methods. – 2019. – Vol.16. – P. 1247-1253. – DOI: 10.1038/s41592-019-0612-7
16. **Deng, J.** ImageNet: A large-scale hierarchical image database / J. Deng, [etc.] // 2009 IEEE Conference on Computer Vision and Pattern Recognition. – 2009. – P. 248-255. – DOI: 10.1109/CVPR.2009.5206848
17. **Ronneberger, O.** U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // International Conference on Medical Image Computing and Computer-Assisted Intervention. – 2015. – P. 234-241. – DOI: 10.1007/978-3-319-24574-4_28
18. **Tomar, N. K.** FANet: a feedback attention network for improved biomedical image segmentation / N.K. Tomar, [etc.] // IEEE Transactions on Neural Networks and Learning Systems. – 2022. – Vol. 34, Issue 11. – P. 9375-9388. – DOI: 10.1109/tnnls.2022.3159394
19. **Long, J.** Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 3431-3440. – DOI: 10.1109/CVPR.2015.7298965
20. **Fan, D.** PraNet: Parallel Reverse Attention Network for Polyp Segmentation / D. Fan, [etc.] // International conference on Medical Image Computing and Computer Assisted Intervention. – 2020. – P. 263-273. – DOI: 10.1007/978-3-030-59725-2_26

DI ZHAO, YI TANG, DMITRY Y. PERTSAU, ALEVTINA B. GOURINOVITCH, DZIANA V. KUPRYIANAVA

LANET FOR MEDICAL IMAGE SEGMENTATION

*Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus*

The paper presents an original LANet model for improving medical image segmentation results based on MobileViT neural network. The developed and integrated Efficient Fusion Attention and Adaptive Feature Fusion blocks improve the quality of feature extraction and reduce data redundancy. The effectiveness of the presented blocks is validated by multiple experiments, including accuracy evaluation on different datasets, based on metrics such as Dice, Precision, Recall, mIoU, model performance evaluation, and ablation study.

Keywords: LANet, Lightweight Attention Network, Efficient Fusion Attention Block, Adaptive Feature Fusion Decoding Block, Attention Mechanism

Ди Чжао, аспирант кафедры информационных технологий в автоматизированных системах, Белорусский государственный университет информатики и радиоэлектроники (БГУИР).

Di Zhao, Postgraduate at the Department of Information Technologies in Automated Systems, Belarusian State University of Informatics and Radioelectronics (BSUIR).

E-mail: 3189124246@qq.com

И Тан, аспирант кафедры информационных технологий в автоматизированных системах, Белорусский государственный университет информатики и радиоэлектроники.

Yi Tang, Postgraduate at the Department of Information Technologies in Automated Systems, Belarusian State University of Informatics and Radioelectronics.

E-mail: tangyijcb@163.com

Перцев Д.Ю., кандидат технических наук, доцент, доцент кафедры электронных вычислительных машин, Белорусский государственный университет информатики и радиоэлектроники.

Pertsau D., PhD, Associate Professor, Associate Professor of Electronic Computing Machines Department, Belarusian State University of Informatics and Radioelectronics.

E-mail: pertsev@bsuir.by

Гуринович А.Б., доцент кафедры информационных технологий автоматизированных систем, Белорусский государственный университет информатики и радиоэлектроники.

Gourinovitch A.B., Assistant Professor at the Department of Information Technologies in Automated Systems, Belarusian State University of Informatics and Radioelectronics.

E-mail: gurinovich@bsuir.by

Куприянова Д.В., старший преподаватель, исследователь, кафедра электронных вычислительных машин, Белорусский государственный университет информатики и радиоэлектроники.

Kupryianava D., Senior Lecturer, Electronic Computing Machines Department, Belarusian State University of Informatics and Radioelectronics.

E-mail: kupriuanova@bsuir.by