

УДК 004.89

Слюсар В.И. (Центральний науково-дослідний інститут озброєння та військової техніки ЗС України)

Слюсарь И.И. (Полтавський державний аграрний університет)

АНАЛИЗ ЭФФЕКТИВНОСТИ МЕХАНИЗМОВ ВНИМАНИЯ В ЗАДАЧЕ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ ЦИФЕРБЛАТОВ АНАЛОГОВЫХ СЧЕТЧИКОВ

Рассмотрены результаты исследования точности различных архитектур нейросетей U-Net с механизмом внимания при решении задачи сегментации изображений циферблатов аналоговых счетчиков. Сделан вывод об отсутствии эффекта увеличения точности за счет интеграции одно- и двухканальных схем внимания, при этом достигнутая точность сегментации составила 86.3 %.

The results of studying the accuracy of various architectures of U-Net neural networks with an attention mechanism in solving the problem of image segmentation of analogue counter dials are considered. It is concluded that there is no effect of increasing accuracy due to the integration of one- and two-channel attention schemes, while the achieved segmentation accuracy was 86.3%.

Важной задачей при реализации концепций Smart City, Smart Home, Industry 5.0 и т.д. является удаленный сбор данных с множества аналоговых средств контроля расхода ресурсов, например, счетчиков потребления электроэнергии, воды, газа, тепла и др. Во многих случаях их замена на современные цифровые приборы является нерентабельной. Немаловажную роль при этом играют большое разнообразие и количество аналоговых средств учета в той или иной инфраструктуре, запреты на внесение изменений в коммуникации, высокая стоимость разработок проектно-технической документации по модернизации оборудования. По этой причине некоторые разработчики идут по пути использования оптических каналов снятия показаний аналоговых счетчиков с последующей передачей изображений их циферблатов на визуальные средства наблюдения. Эффективность такого подхода может быть увеличена путем комбинации технологий Интернета вещей и оптического распознавания данных на основе искусственного интеллекта (IoT + AI).

В качестве возможных примеров подобной интеграции следует указать решения [1]. Однако их массовому внедрению препятствует ряд недостатков, связанных с отсутствием унификации по типам счётчиков, влиянием вибраций при определенной специфике технологических процессов, высокими требованиями к пространственной стабильности изображений и параметрам экспозиции. Кроме того, использование ограниченных возможностей аппаратных средств edge computing вынуждает подстраиваться под минимальные вычислительные ресурсы. При этом возникает необходимость ручной сегментации, точной первоначальной настройки, ручной коррекции считанных данных, и все это – в условиях низкой разрешающей способности изображений (обычно, порядка 28x28 пикселей).

Одним из вариантов решения указанных проблем является использование процедуры предварительной сегментации изображений циферблатов, до того момента, как они поступят на оптическое распознавание. Это позволит снизить объём передаваемой оптической информации, удалить с полученных изображений мешающий фон и сузить поле для распознавания. Варианты такой обработки рассмотрены, например, в [2, 3]. Помимо анализа эффективности различных архитектур нейросетей в [3] было проведено также исследование влияния настроек параметров их типовых слоев (Conv2D MaxPool и др.) в различных комбинациях. Вместе с тем за пределами рассмотрения в [3] остались структуры нейросетей с механизмом внимания, применение которого в задаче сегментации изображений циферблатов счетчиков ранее никем не было изучено.

Целью работы является анализ результатов тестирования новых архитектур нейросетей для сегментации изображений счетчиков с применением механизма внимания.

Для обучения всех предложенных вариантов нейронных структур применительно к рассматриваемой сфере сегментации изображений циферблатов счетчиков использовался общедоступный датасет Water Meters, размещенный на web-портале Kaggle (www.kaggle.com/datasets/tapakah68/yandextoloka-water-meters-dataset).

В качестве базовой архитектуры нейросети была выбрана средняя по размерам U-Net [4] (рис. 1) с количеством параметров 1721730, из которых не подлежали тренировке только 2688 параметров. Особенностью данной структуры U-Net является 4-кратное уменьшение формата матрицы пикселей в процессе обучения. Поэтому размер изображений, поступающих на вход такой нейросети, должен нацело делиться на 4. Вместе с тем, спецификой датасета Water Meters является размер исходных изображений 1000x1778 пикселей, что лишь частично удовлетворяет указанному требованию. Чтобы адаптировать датасет, все его изображения были предварительно пережаты в размеры, кратные 4. В частности, для нового формата изображений была выбрана матрица 224x128 пикселей, как наиболее близкая по своим пропорциям к исходным фотоснимкам. Строго говоря, для изображений со стороной кадра 128 пикселей пересчет с коэффициентом $1778/1000$ дает результат 227,584. В этом случае округление до 224 пикселей должно быть практически незаметным. Следует отметить, что при выборе конкретного формата пережатого изображения помимо максимальной возможности сохранения пропорций исходных фотоснимков датасета также необходимо было учитывать ограничения доступных вычислительных ресурсов, на которых выполняется обучение нейросети. Для максимизации этих ресурсов в рамках исследования использовались возможности графических карт, предоставляемых платным сервисом Google ColabPro+.

Тренировочный датасет содержал 870 снимков, тогда как проверочный сегмент датасета включал 374 изображения. Маски для сегментации циферблатов были заданы черно-белыми. Соответственно процент пространства, занятого черным фоном, составил 98 %, а на белый вырез под цифровое табло счетчика приходилось 2 %. Процесс обучения нейросети U-Net (рис. 1) выполнялся с шагом обучения 0.001 и батчем 32. Время выполнения 80 эпох обучения составило 7 мин в стандартном режиме подключения к Google ColabPro+ с видеокартой A100-SXM4, оснащенной 40 ГБ ОЗУ. В процессе обучения ее ресурсы на некоторых этапах были задействованы более, чем на 87 %. При этом также сервисом Google Colab было предоставлено ОЗУ процессорного модуля объемом 83,48 ГБ и дисковое пространство 166,77 ГБ, хотя их возможности использовались на уровне 6 и 15 процентов соответственно.

Максимальная достигнутая точность обучения U-Net в пределах 400 эпох составила 86.3 %. Комментируя сравнительно невысокий полученный результат следует указать, что датасет Water Meters от Kaggle содержит достаточно много ошибочных масок. Например, некоторым вертикально расположенным циферблатам соответствует горизонтально ориентированная маска. В других случаях при реальном наклоне циферблата влево на маске соответствующий сегмент изображения наклонен вправо. Вполне понятно, что на таких изображениях нейронная сеть дает значительную ошибку, и общая точность сегментации снижается. В подобных случаях исходный датасет следует предварительно отфильтровать от артефактов. Однако, применительно к решаемой задаче, наличие указанного недостатка, напротив, явилось полезным фактором, создавшим стрессовые условия для функционирования анализируемых нейросетей.

На следующем этапе исследования проводились с применением различных по структуре и месту включения в нейронную сеть типа U-Net механизмов внимания. Основная идея применения структур внимания состояла в том, чтобы взвесить матрицу пикселей изображений с помощью особой маски, селективирующей их наиболее значимую область. При этом соответствующие веса подбираются в процессе обучения, максимизируя точность сегментации.

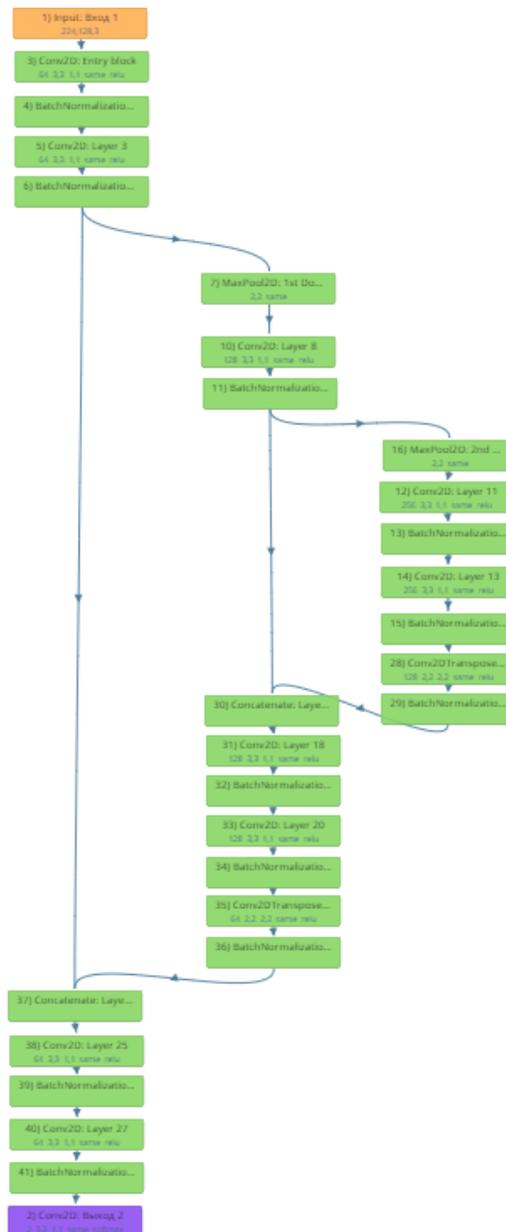


Рис. 1. Базовая архитектура нейросети U-Net в фреймворке “Terra AI”.

Процесс анализа проводился в отношении одно- и двухканальных структур внимания. На рис. 2 представлен один из использовавшихся вариантов одноканального механизма внимания, охватывающий несколько слоев нейросети с пробросом выхода канала внимания вперед по ее структуре. Детализированная на рис. 3 схема механизма внимания содержит на входе сверточный слой Conv2D с одним фильтром, формируемым с помощью единичного весового ядра и такого же шага скольжения. На вход Conv2D поступает изображение формата 112×64 пикселей. Далее следует выравнивающий слой Flatten, выходы которого подключены к полносвязному нейронному слою Dense с функцией активации ReLu. Он имеет количество нейронов, равное количеству пикселей в одном кадре на выходе механизма внимания (1792). Столько же нейронов содержит и последующий слой Dense с функцией активации Softmax. Векторный выход второго слоя Dense далее трансформируется слоем Reshape в изображение формата 52×32 пикселей. Для согласования структуры массивов, поступающих в качестве сомножителей на вход слоя умножения Multiply, на выходе канала внимания в слое Conv2D формируется 256 фильтров, выходы которых подвергаются нормализации в соответствии с процедурой BatchNormalization.

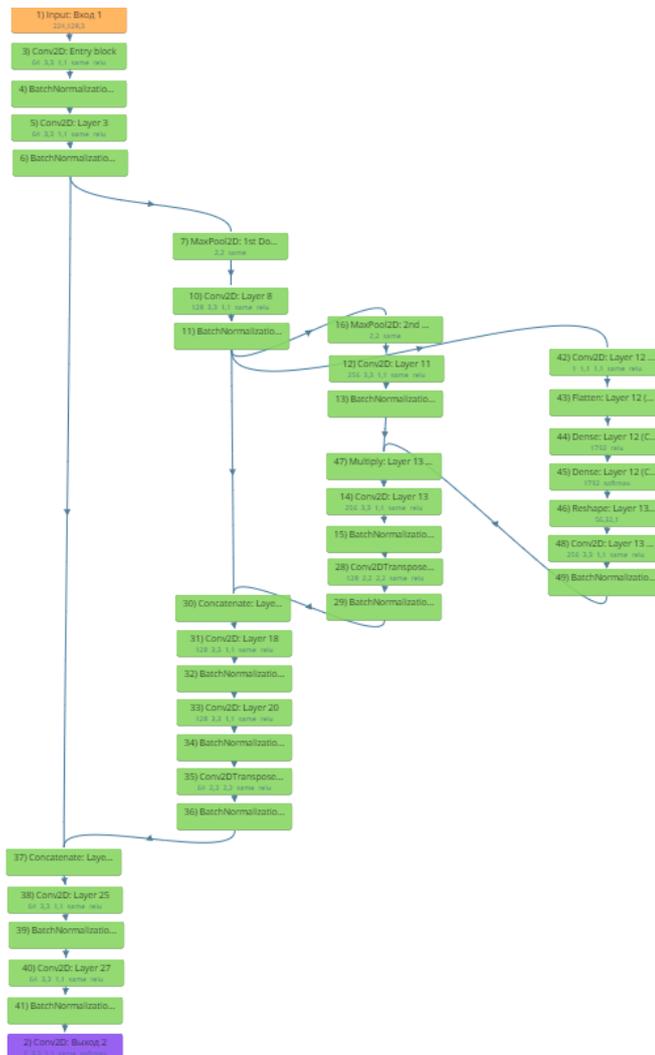


Рис. 2. U-Net с механизмом внимания, охватывающим 3 слоя нейросети.

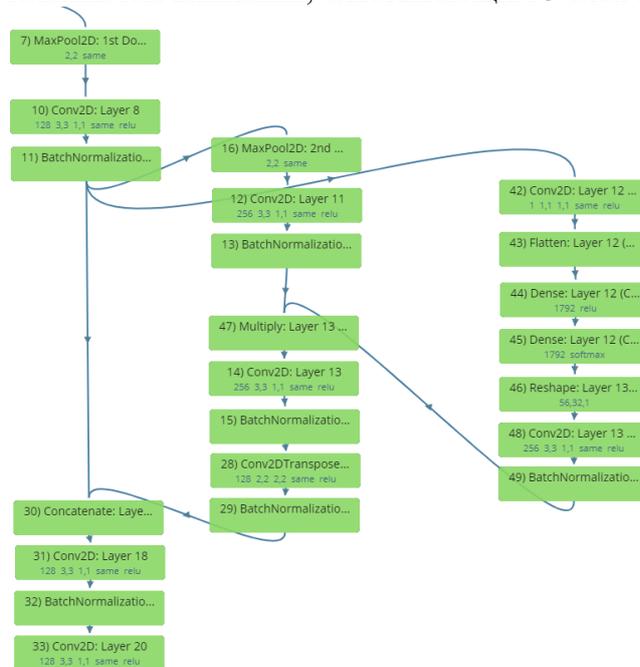


Рис. 3. Структура одноканального механизма внимания на рис. 2.

Обучение нейросети рассмотренной структуры позволило получить точность 85.1 % в пределах 200 эпох. При этом наблюдалось затягивание начала процесса

ухудшению точности сегментации, которая снизилась до величины 75.3 %. При этом также заметно ухудшилась и скорость обучения.

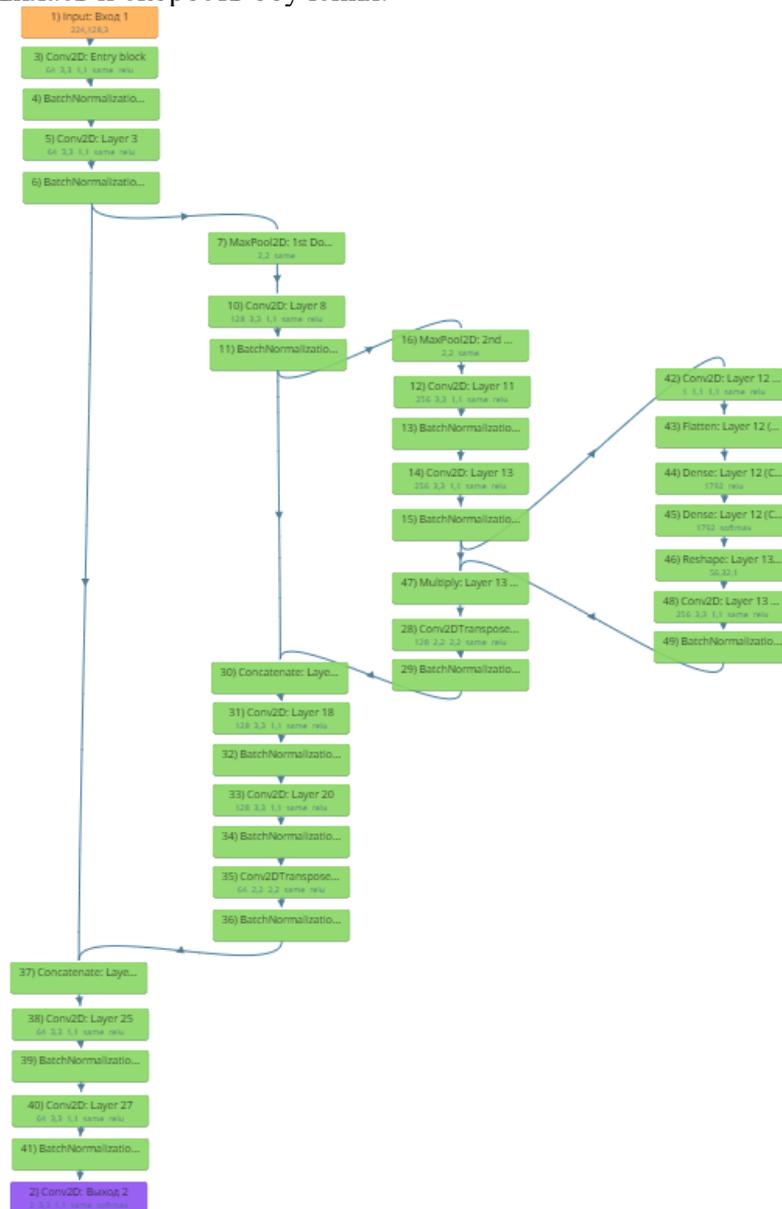


Рис. 5. Альтернативная архитектура U-Net с одноканальным механизмом внимания, подключенным в межслойный интервал.

Данный результат позволил сделать однозначный вывод, что применительно к рассматриваемому датасету одноканальный (однокаскадный) механизм внимания является предпочтительным решением. Также более выигрышным вариантом является использование в схеме внимания последовательно двух слоев Dense вместо одного. Такой принцип реализации механизма внимания на спарке слоев Dense дает преимущества в точности сегментации и по сравнению с сугубо сверточной схемой, содержащей несколько последовательно включенных сверточных слоев и слоев нормализации.

Естественно, в рамках проведенных исследований невозможно было охватить все варианты реализации механизма внимания. Вполне вероятно, что некоторые из них могут оказаться более эффективными. Проверка данного предположения является целью дальнейших исследований.

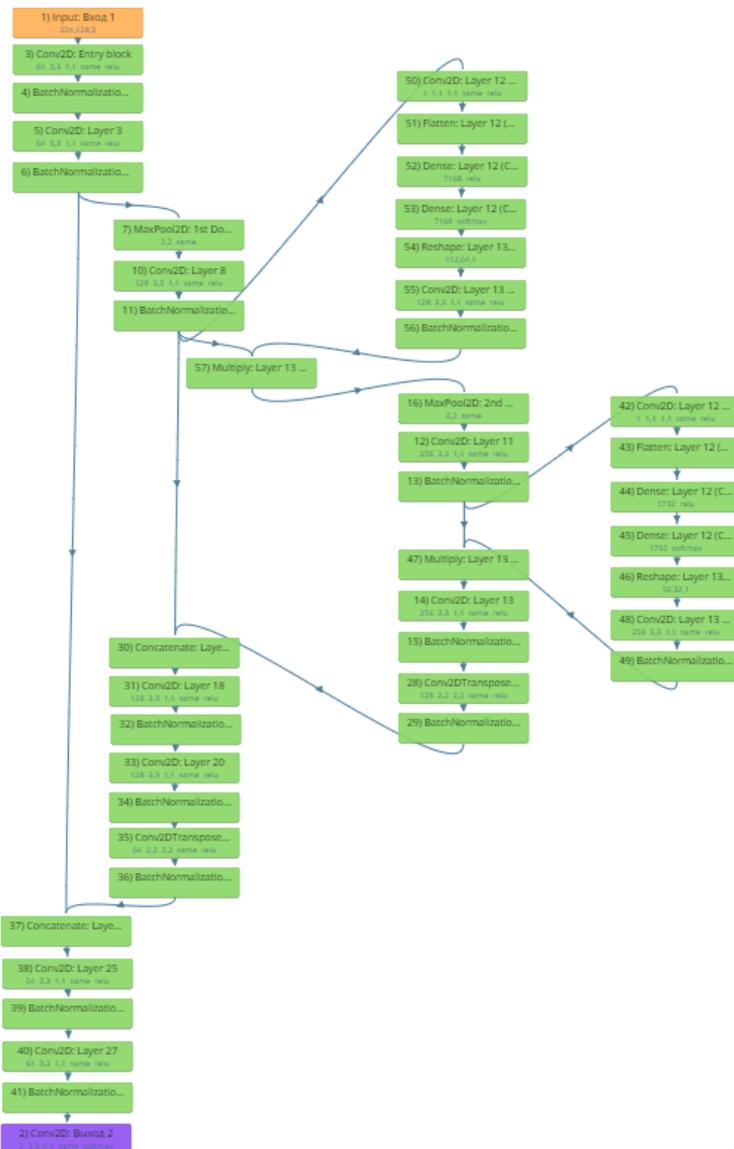


Рис. 6. U-Net с двухканальным механизмом внимания.

ВЫВОДЫ

Полученные результаты позволили сделать вывод, что применительно к датасету Water Meters механизм внимания, интегрированный в U-Net средней сложности, не позволил улучшить точность сегментации по сравнению с базовой архитектурой. В проведенных экспериментах удалось получить точность сегментации 86,3 %, как в случае использования механизма внимания, так и без него. Однако, не исключено, что использование более сложных схем внимания в сочетании со структурным разнообразием архитектур U-Net позволит преодолеть указанный паритет в пользу нейросетей с интегрированным механизмом внимания.

Литература

1. Analog meters in the digital enterprise: change or integrate? URL: <https://habr.com/ru/company/lanit/blog/676240/>.
2. Vadym Slyusar, Mykhailo Protsenko, Anton Chernukha, Vasyl Melkin, Olena Petrova, Mikhail Kravtsov, Svitlana Velma, Nataliia Kosenko, Olga Sydorenko, Maksym Sobol. Improving a neural network model for semantic segmentation of images of monitored objects in aerial photographs. // Eastern-European Journal of Enterprise Technologies.- № 6/2 (114). – 2021. - Pp. 86 – 95. DOI: 10.15587/1729-4061.2021.248390.
3. Vadym Slyusar, Ihor Sliusar, Nataliia Bihun, and Volodymyr Piliuhin. Segmentation of analogue meter readings using neural networks. // 4th International Workshop on Modern Machine Learning Technologies and Data Science MOMLET&DS2022, Lviv, Ukraine, 23 -25 Nov. 2022. – 11 p.
4. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 234–241.