ИНФОРМАТИКА И УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМАХ

УДК 544.454; 536.46; 614.841.1

В.С. Бочков[1], Л.Ю. Катаева[1,2], Д.А. Масленников[1], И.В. Каспаров[2]

ПРИМЕНЕНИЕ АРХИТЕКТУРЫ ГЛУБОКОГО ОБУЧЕНИЯ U-NET ДЛЯ РЕШЕНИЯ ЗАДАЧИ ВЫДЕЛЕНИЯ ВЫСОКОТЕМПЕРАТУРНЫХ ЗОН ПОЖАРА НА ВИДЕО

Нижегородский государственный технический университет им. Р.Е. Алексеева[1]
Самарский государственный университет путей сообщения[2]

Представлены результаты применения методов глубокого обучения к задаче сегментации объектов пламени на видео. Приведено сравнение моделей различных U-Net архитектур, предложена модернизация исходной архитектуры путем анализа анимированных последовательностей распространения огня. Обоснована эффективность использования данной методологии для выявления конфигурации пламени на видео. Представленная базовая модель на основе использования 11-слойной архитектуры энкодера VGG11 может быть улучшена с применением более мощных образцов, например, SE-ResNet 101. Модель может быть использована на микрокомпьютерах с графическим чипом и применима в робототехнике, в частности, при создании роботизированных средств контроля и устранения очагов возгорания. Использование RGB-видеопотока открывает возможности экономически выгодных решений по сравнению со многими современными средствами зонального слежения за ситуацией.

Впервые решена задача точной сегментации пламени с видеопотока в режиме реального времени на основе метода UNet. Решена задача сегментации как одноклассовой сегментации пламени (есть пламя в пикселе/нет пламени в пикселе), так и трехклассовой, разделенной по цвету (красное/оранжевое/желтое пламя). Впервые использован подход к детекции объектов с использованием анимированных последовательностей изображений, который показал существенный прирост точности.

Ключевые слова: глубокое обучение, сегментация, Sorensen-Dice, Jaccard, U-Net, аугментации.

Введение

Пожары являются наиболее распространенным типом природных и техногенных катастроф, наносящих вред экономике и здоровью человека. Наиболее частой причиной возникновения лесных возгораний, наряду с неосторожным обращением с огнем, является несвоевременное обнаружение проблемы и запоздалое реагирование, что, в свою очередь, обусловлено отсутствием средств зонального слежения за ситуацией. В связи с этим остро стоит проблема создания роботизированных средств контроля и устранения очагов возгорания. Среди современных средств контроля за обстановкой можно выделить использование бесконтактных ИК-датчиков тепла или тепловизоров. Однако их цена достаточно высока для широкого внедрения технологии и обеспечения слежения за обстановкой в лесных массивах с использованием беспилотных летательных аппаратов.

В настоящей статье представлены результаты применения к задаче сегментации объектов пламени на видео методов глубокого обучения, которые ранее доказали свою эффективность в решении задач компьютерного зрения [1, 2]. Базируясь на природе нелинейных преобразований, представляемых в виде композиции нейронных слоев, модель может извлекать нетривиальные паттерны обучаемых явлений. Несмотря на то, что сверточные нейронные сети существуют достаточно долгое время, их применение в реальных задачах было ограничено

требованиями к гигантским размерам датасетов изображений, сборка которых является достаточно трудоемкой. Это обусловлено использованием полносвязных слоев нейронной сети в конце энкодера, кратно увеличивающих количество параметров сети, необходимых для тренировки. С введением аугментаций (преобразований) входных изображений их количество уменьшилось, но, тем не менее, осталось значительным. Поскольку полносвязные слои образуют основное количество всех параметров сети, было предложено создавать новые архитектуры нейросетей из базовых путем отсечения последних полносвязных слоев [3]. Такие архитектуры получили название полностью-сверточных сетей, где на выходе мы имеем малоразмерную матрицу сигнала присутствия объектов на изображении вместо вектора объектов, из которого находится наиболее подходящий в задаче классификации. Установлено, что данная маска может быть использована в задаче сегментации изображений с локализацией объектов на нем путем увеличения размера маски (upsampling). Первая архитектура для задачи сегментация называлась FPN8 [3], где матрицу последнего слоя отсекали по пороговым значениям и увеличивали кратно в 8 раз до размеров исходного изображения, получая маску сигнала на нем. Следующим методом за данной достаточно грубой аппроксимацией стал SegNet, суть которого заключалась не в однократном увеличении маски сигнала в 8 раз, а увеличением его в 4 раза, вдвое с последующим проходом сверточного слоя, составляя декодер модели [4]. Этот метод значительно улучшил аппроксимацию объектов на видео по сравнению с FPN8, но по-прежнему выдавал не лучшее маскирование, руководствуясь на проходе развертки только исходными данными последней маски энкодера. U-Net [5] является модернизацией модели SegNet, где особенностью является использование пропускных коннекторов между слоями одного размера (рис. 1). В рамках представленной вычислительной модели данной архитектуры слева продемонстрирован энкодер данных, осуществляющий их сжатие, по правую сторону происходит раскрытие слоев с использованием операций двукратного увеличения размера маски и конкатенации с результатом, полученным на энкодерном слое того же уровня. Данный проход называется декодированием.

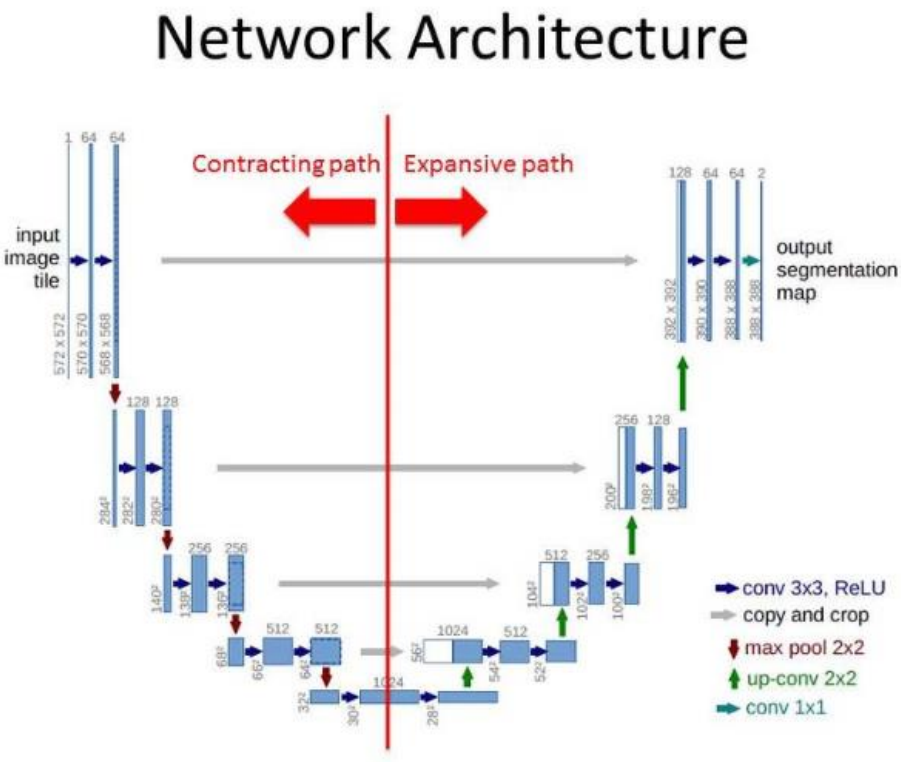

Рис. 1. U-Net архитектура нейронной сети

Обоснованием использования данного типа коннекторов является сохранение локации максимального сигнала обнаружения нужного объекта на определенном слое для уточнения его координат на выходе сети. В настоящей работе описано применение данного метода.

Данные исследования

Поскольку полностью-сверточные сети не нуждаются в громадном объеме данных, U-Net архитектуры могут быть применены для решения реальных коммерческих задач, и сбор данных обучения не занимает большого времени. Для сравнения: задачи классификации базируются на тренировке нескольких миллионов изображений (Mega-face датасет объемом ~ 1млн изображений, ImageNet – ~ 14 млн), в то время как в решении задач сегментации медицинских изображений достаточно неплохие результаты были достигнуты уже при использовании 20-30 изображений [5]. Малое количество изображений обусловлено тем, что в решении используются различные аугментации (преобразования), в том числе, искажения исходных изображений (клеточная структура на изображении подвержена деформациям, поэтому использование данных видов аугментаций обосновано), которые не могут быть применимы в задаче обнаружения огня на реальных объектах. Однако разворот изображения по горизонтали позволяет увеличить размер датасета как минимум вдвое, и это минимальная аугментация, которая использовалась в начале исследований. В задаче сегментации пламени применялся датасет размером в 200 изображений аугментированных до 6 000 сэмплов.

Метрики точности и функции ошибки

Поскольку задача сегментации изображений отличается от классификации, для ее решения необходимо использовать метрики попиксельной точности и корректности модели, а также функции ошибки, градиенты которой указывали на изменение коэффициентов в сторону их минимизации. Первой идеей тренировки и валидации модели было использование стандартной категориальной кросс-энтропии на пиксельном уровне: в модели было N^2 классификаторов, функции ошибки которых суммировались, и находилось среднее значение, от которого вычислялся градиент для корректировки коэффициентов. В задачах сегментации используются две основные метрики. Первая – мягкий Jaccard index [6] (пересечение через объединение), описываемый следующей формулой (1):

$$J_s(c) = \frac{P_c \cap Y}{P_c \cup Y} = \frac{|P_c \cap Y|}{|P_c| + |Y| - |P_c \cap Y|} = \frac{\sum_i p_{c_i} y_i}{\sum_i p_{c_i} + \sum_i y_i - \sum_i p_{c_i} y_i}. \quad (1)$$

Второй метрикой является мягкий Sorensen-Dice [7, 8] (2):

$$D_s(c) = \frac{2|P_c \cap Y|}{|P_c| + |Y|} = \frac{2\sum_i p_{c_i} y_i}{\sum_i p_{c_i} + \sum_i y_i}. \quad (2)$$

Здесь p_{c_i}, – вероятность предсказания моделью нахождения объекта c в пикселе, i, y_i – бинарные значения актуального объекта в пикселе. Для осуществления сегментации изображений выходная матрица должна быть преобразована по сигмойде (3):

$$s(n) = \frac{1}{1 + e^{-p_n}}, \quad (3)$$

где n – номер класса; а также дискретизирована по порогу (4):

$$S = \begin{cases} n, \max\limits_{n \in N}(s_i(n)) \geq thr \\ 0, \max\limits_{n \in N}(s_i(n)) < thr \end{cases} \quad (4)$$

где N – количество классов в задаче сегментации.

Для валидации модели в процессе поиска наилучшего порога обнаружения используют метрики Dice и Jaccard, в которые на вход приходят бинаризованные по порогу вероятности нахождения пикселей. В данной работе поиск лучшего порога производится путем полного

перебора точек отрезка [-6, +6] с шагом 1e-3, в котором определена функция сигмойды. Установлено, что наилучшие результаты обнаружения получаются при использовании порога в области от нуля до единицы, что показано на рис. 2.

На оси абсцисс отображено значение порога обнаружения, на оси ординат – показатель точности модели при использовании данного порога. Отображены значения порога, при котором достигаются максимальные показатели.

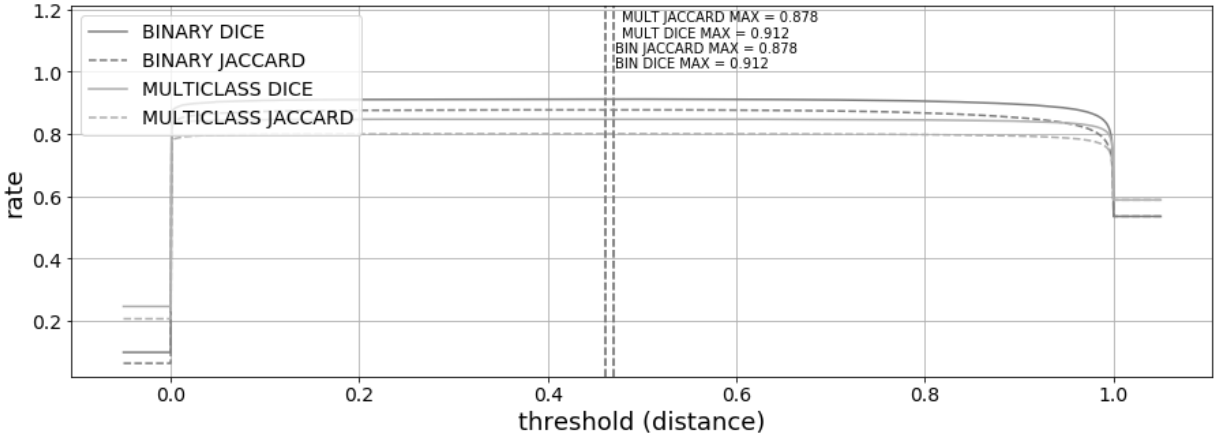

Рис. 2. График распределения точности обнаружения огня относительно выбора порога обнаружения

В работе используется функция ошибки, полученная в результате логарифмирования функции мягких метрик Dice & Jaccard и ее композиции с функцией бинарной кроссэнтропии (5):

$$BCE = -\frac{1}{N}\sum_n ys(n) + (1-y)\ln(1-s(n)),$$
$$L = \frac{BCE - \ln(J_S)}{2}, \quad (5)$$
$$L = \frac{BCE - \ln(D_S)}{2}.$$

Бинарная и многоклассовая сегментация пламени для выявления зон

Бинарная сегментация представляет собой извлечение всех пикселей пламени без их разделимости. Многоклассовая сегментация подразумевает разделение пламени на пламя красного, оранжевого и желтого цвета на видео. Поскольку цвет пламени напрямую зависит от его температуры и материала горения [9], имеет смысл извлекать данную дополнительную информацию для последующего анализа уязвимых точек пламени. Результаты тренировки бинарных задач сегментации, в основе которых лежат функции мягкого Dice и Jaccard-коэффициентов, показаны на рис. 3. В приведенном графике ступенчато обозначены показатели точности бинарного Dice и Jaccard-коэффициентов в среднем. Тонкими вертикальными линиями отмечены показатели среднеквадратичного отклонения метрик на множестве тестовых данных. Данные показатели остаются достаточно большими, что сигнализирует о склонности модели к переобучению на тренировочном датасете, и отсутствием генерализации на другие данные. Показатели, улучшенные на 10 % точности, представлены для модели, натренированной на основе мягкого Jaccard-индекса. Результаты мультиклассовой сегментации показаны на рис. 4.

В данном случае мы наблюдаем показатели 4 метрик точности. Бинарные показатели хуже, чем у бинарных моделей, однако разница между ними уменьшается с введением аугментаций данных в тренировочный датасет. Как и в случае с бинарными моделями, многоклассовые, натренированные на мягком Jaccard индексе, показывают лучшие результаты.

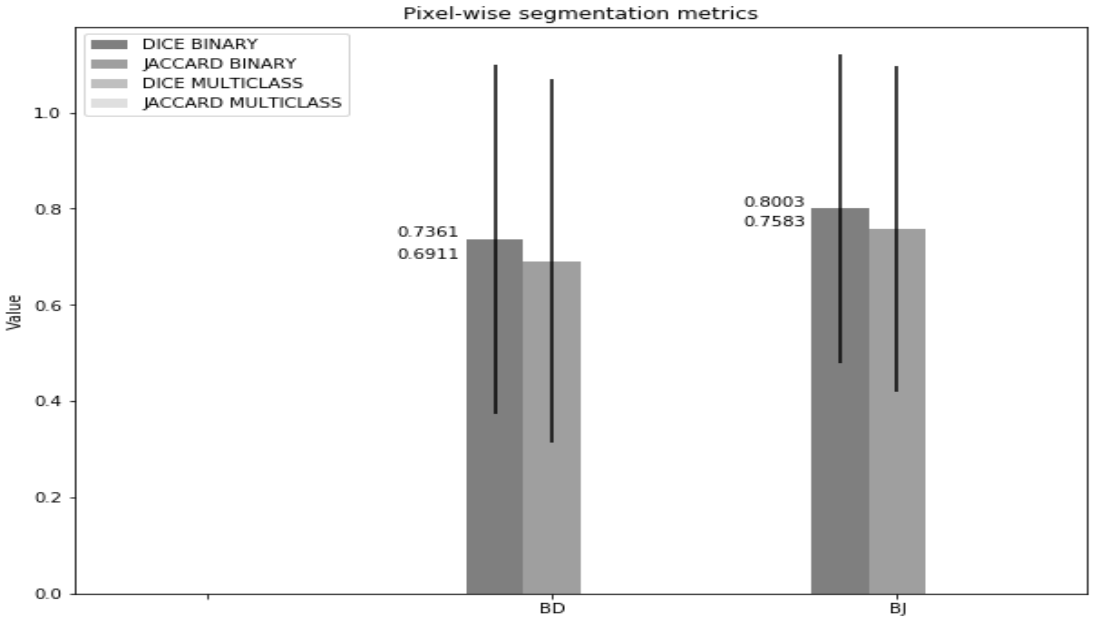

Рис. 3. Графики точности бинарных моделей сегментации

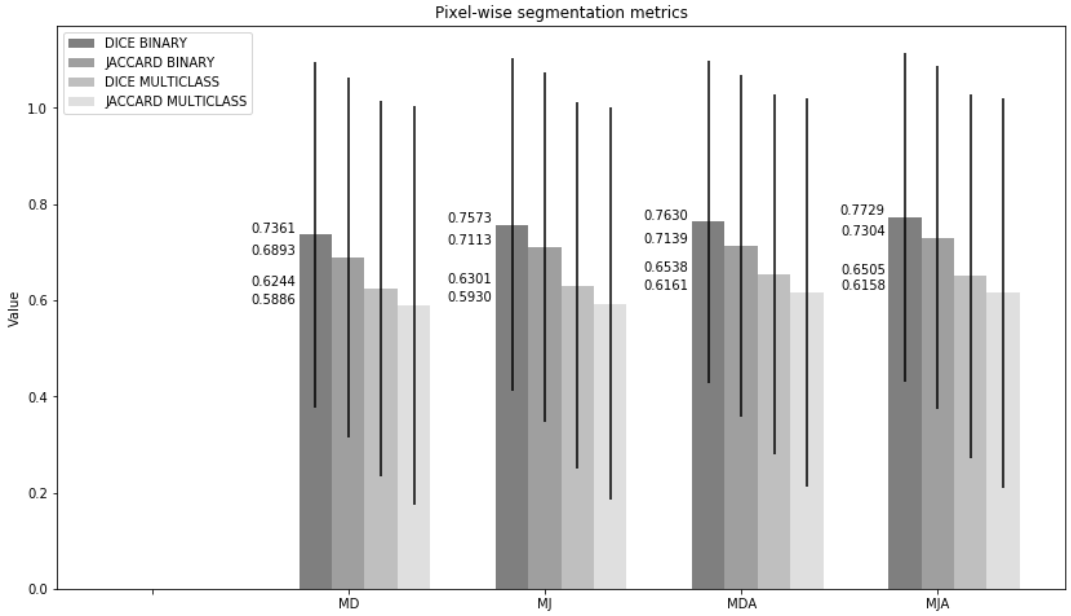

Рис. 4. Показатели точности моделей мультиклассовой сегментации

При визуальном анализе видеофайлов горения покадрово установлено, что пламя является достаточно динамичным объектом. Учитывая этот фактор, целесообразно проводить анализ объектов, используя непрерывные анимированные данные, предшествующие последнему кадру, на котором осуществляется маскирование пламени. Как показано на рис. 5, использование анимаций в качестве входных данных существенно увеличивает точность обнаружения и уменьшает среднеквадратичное отклонение. Одним из важных факторов улучшения показателей является, в частности, использование во время тренировки реверсивных анимаций горения, что позволяет увеличить размер датасета в два раза.

Результат распознавания четырехкадровой модели представлен на рис. 6, 7. Для четкой визуализации в непомеченных пламенем участках произведена перемена местами красного и синего канала изображения.

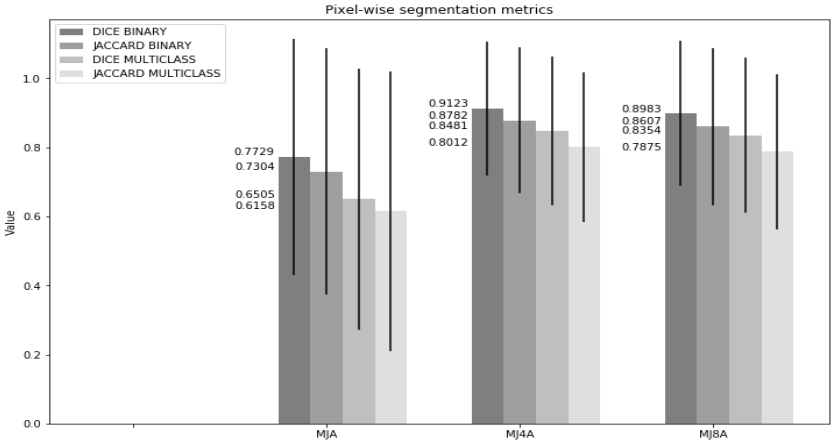

Рис. 5. Графики точности четырех- и восьмикадровых анимированных моделей в сравнении с однокадровой моделью

Рис. 6. Результат обнаружения пламени высокотемпературных зон пожара моделью, базирующейся на четырехкадровом анализе

Рис. 7. Результат обнаружения пламени высокотемпературных зон пожара моделью, базирующейся на четырехкадровом анализе

Заключение

Полученные результаты свидетельствуют об эффективности использования методов глубокого обучения для выявления конфигурации пламени на видео. Точность 90 % достигнута благодаря анализу анимированных фрагментов. Модель является базовой, на основе использования 11-слойной архитектуры энкодера VGG11, и может быть улучшена с применением более мощных моделей, например, SE-ResNet 101. Продемонстрированы отличные показатели генерализации, однако для промышленного внедрения данных моделей необходимо набирать большее количество данных и проводить дополнительные итерации тренировки, поскольку на момент анализа различных видео с квадрокоптеров, либо с экшн-камеры сотрудников МЧС, возникают различного рода ошибки ложного распознавания. Они могут быть нивелированы путем выявления таких участков видео и внедрением их в множество тренировочных данных. Модель в режиме тестирования расходует 2 Гб видеопамяти и исполняется за 1 мс, поэтому может быть использована на микрокомпьютерах с графическим чипом, например, Nvidia Jetson Nano, применяемых в робототехнике.

Библиографический список

1. **Krizhevsky, A.** Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // NIPS. – 2012. – P. 1106-1114.
2. **LeCun, Y.** Backpropagation applied to handwritten zip code recognition / Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel // Neural Computation. – 1989. – № 1(4). – P. 541–551.
3. **Long, J.** Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // arXiv:1411.4038 [cs.CV]. – 2014.
4. **Badrinarayanan, V.** SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling / V. Badrinarayanan, A. Handa, R. Cipolla // arXiv:1505.07293 [cs.CV]. – 2015.
5. **Ronneberger, O.** U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // arXiv:1505.04597 [cs.CV]. – 2015.
6. **Jaccard, P.** ´Etude comparative de la distribution florale dans une portion des Alpes et des Jura / P. Jaccard // Bulletin de la Soci´ et´e Vaudoise des Sciences Naturelles. – 37:547–579. – 1901.
7. **Sørensen, T**. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons / T. Sørensen // Kongelige Danske Videnskabernes Selskab. – 1948. – № 5 (4).
8. **Dice, Lee R**. Measures of the Amount of Ecologic Association Between Species / Lee R. Dice // Ecology. –1945. – № 26 (3): 297–302. doi:10.2307/1932409. JSTOR 1932409.
9. **Бочков, В.С.** Алгоритм поиска уязвимых зон пожара с применением анализа видеопотока / В.С. Бочков, Л.Ю. Катаева, Д.А. Масленников // Сборник тезисов XXIX международной научно-практическая конференция, посвященная 80-летию ФГБУ ВНИИПО МЧС России. – 2017. – С. 395-400.

Дата поступления
в редакцию: 10.06.2019

V.S. Bochkov[1], L.Yu. Kataeva[1,2], D.A. Maslennikov[1], I.V. Kasparov[2]

APPLICATION OF U-NET DEEP LEARNING ACHITECTURE FOR FIRE SEGMENTATION ON VIDEO

Nizhny Novgorod state technical university n.a. R.E. Alekseev[1]
Samara state university of railway transport[2]

Purpose: multiclass image segmentation of fires.
Methodology: we use U-Net deep learning architecture over one-image and multi-image animations.
Value: the represented in paper result of fire-segmentation over animation is robust. Got big accuracy values with little mean square error. It can be applied to further segmentation of best plases to suppress the fire.
Research implications: the approach of using U-Net architecture over animation data can be implemented in big variety of human-assistance devices and automated water-cannons.

Keywords: deep-learning, segmentation, Sorensen-Dice, Jaccard, U-Net, augmentations.