

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ОБРАБОТКА ДАННЫХ

Научная статья

УДК 519.25

DOI: 10.18101/2304-5728-2023-1-37-46

АРХИТЕКТУРЫ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ И ИХ ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ДЛЯ БЕСПИЛОТНОЙ ПОСАДКИ ЛЕТАТЕЛЬНОГО АППАРАТА

© Кузьмин Олег Викторович

доктор физико-математических наук, профессор,
заведующий кафедрой теории вероятностей и дискретной математики,
Иркутский государственный университет
Россия, 664003, г. Иркутск, ул. Карла Маркса, 1
quzminov@mail.ru

© Лавлинский Максим Викторович

аспирант,
Иркутский государственный университет
Россия, 664003, г. Иркутск, ул. Карла Маркса, 1
lavlinskimv@mail.ru

Аннотация. В данной работе рассматриваются архитектуры сегментации изображений и их практическое применение для беспилотной посадки, примеры реализованных систем доставки товаров дронами. Выбрана и реализована оптимальная посадка дрона с использованием сегментации изображения нейронной сетью. Выбран и модифицирован набор данных для обучения, путем отзеркаливания изображения и разворота изображения. Создана дополнительная выборка данных с использованием приложения, симулирующего полёт дрона в настраиваемом окружении. Найден способ быстрой разметки видеозаписи приземления дрона.

Представлены результаты тестирования нейронных сетей различных архитектур для выявления оптимальной для решаемой задачи. Были применены две разные метрики точности нейронной сети для получения достоверных данных о распознавании. Были выбраны наиболее точно узнаваемые классы окружения. Для этого было создано три варианта нейронных сетей с разным набором классов. Результаты были проанализированы и выбраны наиболее узнаваемые классы.

Создана финальная нейронная сеть на основе архитектуры LinkNet. Исходя из высоты дрона выбрана зона приземления и зона для последующей сегментации. Найдена оптимальная высота полета над поверхностью для начала процедуры приземления.

Ключевые слова: нейронная сеть, сегментация изображения, беспилотный дрон, безопасная посадка, дистанционное управление.

Для цитирования

Кузьмин О. В., Лавлинский М. В. Архитектуры сегментации изображений и их практическое применение для беспилотной посадки летательного аппарата // Вестник Бурятского государственного университета. Математика, информатика. 2023. № 1. С. 37–46.

Введение

Поиск мест для безопасной посадки летательного аппарата может быть использован во многих сферах: от бизнеса по доставке продуктов до мониторинга лесных массивов и оперативных работ [16, 17, 18, 19, 20]. Также определение мест безопасной посадки может помочь в управлении дронами операторами, как в системе доставки *Mapna*, в которой оператор лишь подтверждает зону посадки, выбранную нейросетью. Это позволяет свести случаи крушений при посадке к минимуму и существенно увеличить эффективность работы операторов.

На определение местности, опасной и безопасной для посадки, влияет множество факторов от отблеска на мокрой поверхности до проезжающей машины. Самыми сложными факторами для выявления являются движущиеся объекты, которые вскоре могут повлиять на безопасность посадки и уже посаженный дрон.

Один из способов осуществления безопасной посадки использует анализ топологических особенностей местности, но это, как правило, не даёт удовлетворительного результата и не у любой местности есть уже созданная топологическая карта. В работе рассматриваются системы доставки, использующие GPS-данные в комбинации с нейронными сетями для посадки беспилотных и пилотируемых дронов. Такой способ часто требует дополнительных наблюдений. Например, идущий пешеход не будет найден нейронной сетью, поскольку в момент определения места для посадки не будет находиться на месте приземления.

Как правило, для доставки используются квадрокоптеры, поскольку для посадки им необходимо пространство, ненамного большее самого дрона. Но они имеют недостатки: большое потребление энергии и высокая мощность моторов. Для поддержания аппарата в воздухе требуется заметно больше ёмкости батареи, чем у аналогичных по дальности полёта летательных аппаратов на основе конструкции самолёта, что увеличивает стоимость дрона и, следовательно, потери при его крушении.

Установка тренированной CNN [1] нейронной сети на дрон для распознавания мест приземления резко увеличивает стоимость дрона, энергопотребление и вес, но убирает необходимость в передаче данных на сервер для расчётов и отправки данных обратно на дрон, что является главной целью исследования. Поэтому требуется максимально возможно уменьшить время на обработку кадра. Уменьшение размера и сложности сети также снижает точность предсказания и время обработки кадра, но при обработке изображения появляются новые возможности. Во-первых, можно определить оптимальное количество кадров в секунду. Обработка 120 кадров в секунду слишком затратна и принципиально увеличивает

точность. Обработка видео по 10 кадров в секунду займет значительно меньше времени. Во-вторых, уменьшить разрешение картинки. В-третьих, обрезать изображение в зависимости от высоты полёта, поскольку во время полёта камера будет захватывать изображение, где предполагаемое место для посадки отображено лишь небольшим набором пикселей.

Цель данной работы — создание ориентированного на изображения метода безопасной посадки беспилотного дрона, имеющего достаточную точность и минимально возможное потребление энергии и вычислительной мощности.

1 Данные для обучения и тестирования

Основной источник данных для обучения — набор данных воздушной сегментации Градского технологического университета [2]. Набор данных содержит 400 изображений, снятых на высоте 15 метров, размера 6000 на 4000 пикселей и маски этих изображений. Маски содержат классы: земля, трава, гравий, вода, камни, растительность, плитка, крыши, окна, двери, стены домов, заборы, столбы, люди, собаки, машины, велосипеды, мертвые деревья, дорожные препятствия, маркер (специальный символ, который может обозначать подготовленное место посадки). Не все классы имеют значение при определении места посадки, поэтому для тестирования и обучения нейронных сетей некоторые классы ввиду их малой информативности будут преобразованы в один. Для увеличения количества данных применено зеркальное и вертикальное изображение.

Для тестирования работы нейронной сети на видео информации использован второй источник данных. Видео были взяты из игры, стимулирующей полёт дрона AI DRONE SIMULATOR. Данная игра позволяет записывать видео полёта дрона и манипулировать положением и углом наклона камеры. В игре существует набор окружений и простой редактор карт. На рисунке 1 изображение из игры воссоздает окружение реального мира. В статье “Using Video Games to Train Computer Vision Models” [3] приведены примеры подобного использования игровых окружений для сбора данных и тренировки нейросетей.

При использовании виртуального окружения, которое изначально не было предназначено для сбора данных, выявлено множество недостатков. Основные из них — невозможность получения данных о положении и скорости дрона, сложность в его управлении и стабилизации, невозможность автоматического создания маски снимка исходя из объектов окружения. Для съемки удовлетворительного видео приземления была отключена возможность управления дроном, что позволило его стабилизировать. Оставалась проблема с определением высоты дрона на каждом кадре видео: чтобы высчитать высоту, была увеличена скорость дрона до максимально возможной — 20 км/ч. Поскольку съемка производилась с частотой 30 кадров в секунду, можно вычислить, что за каждый кадр дрон пролетел около 19 сантиметров. Зная время приземления, можно опреде-

лить высоту дрона на каждом кадре. 20 км/ч — скорость, большая для реального приземления, но позволила собрать данные для последующего анализа работы нейронной сети.

Последней проблемой стало создание маски видео для сравнения её с данными, полученными нейронной сетью. Создание маски каждого кадра видео вручную потребует разметки 900 кадров (6 видео по 5 секунд с частотой 30 кадров в секунду), что практически невозможно. Для обхода этой проблемы создана маска первого кадра изображения, поскольку видео посадки было снято неподвижно относительно осей OY и OX (изменялась только высота, то есть координата по оси OZ), первый кадр содержит в себе все последующие. Зная высоту, можно обрезать размер маски первого кадра и получить маску нужного кадра, что полностью избавляет от необходимости разметки каждого кадра.

2 Выбор и тренировка нейронных сетей

Самым важным фактором в выборе нейронной сети стала скорость обработки видео. Для решаемой задачи сегментации изображения существует небольшое количество действенных архитектур, а именно PSPNet [4], LinkNet [5], Unet [6], FPN [7], основанных на свёрточных нейронных сетях [8] и архитектуре автокодировщиков [9]. Полная сегментация изображения очень долгий процесс, требующий большого количества параметров сети. Поэтому были созданы 4 нейронные сети примерно с одинаковым количеством параметров, проанализированы результаты их обучения и скорость обработки одного кадра.

Тренировка проходила с двумя классами. Безопасные места для посадки: трава, земля, гравий, плитка. Опасные — вода, камни, растительность, крыши, окна, двери, стены домов, заборы, столбы, люди, собаки, машины, велосипеды, мертвые деревья, дорожные препятствия. После обработки данных нейронной сетью проведена оценка вероятности предсказания класса с использованием функции активации выходного слоя ‘softmax’ [10]. Использованная функция ошибки — ‘categorical_crossentropy’ [11]. Метрики точности: ‘IOU’ — Intersection-OverUnion [12], F1-score [13].

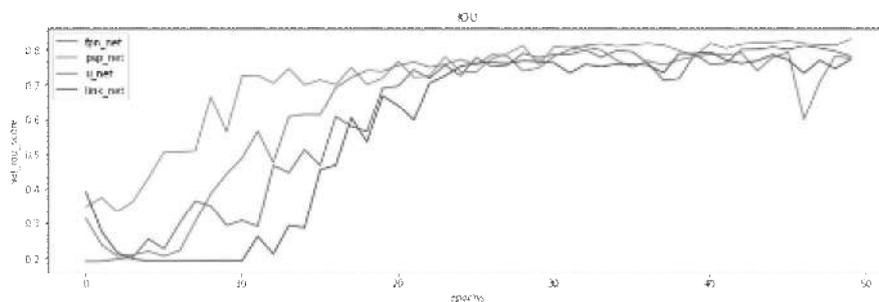


Рис. 1. График Intersection-OverUnion на каждой эпохе обучения

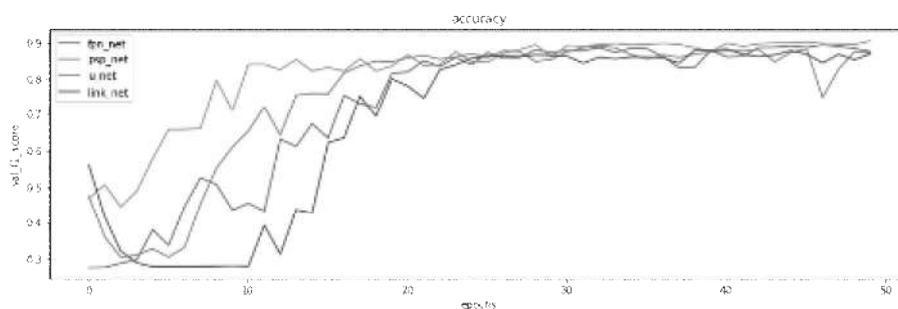


Рис. 2. График F1-score на каждой эпохе обучения

Таблица 1. Результаты обучения

Архитектура	Максимальная точность (iou)	Время на обработку одного кадра	Время на обучение	Количество параметров
Unet	0.8097	1130 мс	4:52:28	20,325,427
PSPnet	0.8323	1601 мс	7:47:33	19,345,115
LINKnet	0.7961	803 мс	4:52:26	20,325,717
FPN	0.8145	1550 мс	8:25:31	17,577,157

Из-за ограниченности ресурсов тренировка нейронных сетей проходила на уменьшенном сете данных, поэтому не удастся точно определить, какая архитектура показала лучший результат, поскольку из-за маленькой выборки очень страдает точность. Но каждая из нейронных сетей показала удовлетворительный результат (более 0.75), поэтому была выбрана самая быстрая архитектура (LinkNet) и в последующих тренировках были использованы нейронные сети только с этой архитектурой.

3 Выбор классов

Помимо классов, определяемых как безопасные (трава, земля, гравий, плитка), существует несколько классов, в определении которых нейронная сеть может ошибиться, — вода и крыши. Для того чтобы определить самые точно определяемые классы, было создано три варианта нейронных сетей. Первая из них ищет траву, плитку и остальные классы, вторая — траву, плитку, гравий, землю и остальные классы, третья — траву, плитку, воду, крыши и остальные классы.

Как можно заметить из результатов обучения (табл. 2), в выборке существуют классы, очень сложные в определении, что вызвано несбалансированностью набора данных для тренировки и возможной схожестью с другими классами. Поэтому даже если возможна посадка на землю или гравий, выбирать их как безопасные для посадки классы невыгодно, поскольку это сильно влияет на общую точность нейронной сети, даже если при этом уменьшится количество возможных посадочных мест.

Таблица 2. Результаты обучения

Нейронная сеть	Плитка	Трава	Земля	Гравий	Вода	Крыша	Другое	Среднее
сеть 1	0.9033	0.9335	-nan	-nan	0.1730	0.6588	0.8163	0.694
сеть 2	0.8267	0.9483	0.03	0.112	-nan	-nan	0.7376	0.502
сеть 3	0.7890	0.9721	-nan	-nan	-nan	-nan	0.7751	0.845

4 Финальная нейронная сеть

Архитектура LinkNet — бинарная классификация на два класса. Для более быстрой тренировки был выбран backbone [14] mobilenet [15], уменьшено количество параметров до 4,546,210, что позволило ускорить обработку кадра до 93 мс. Функция ошибки — ‘categorical_crossentropy’, метрика — ‘Intersection-OverUnion’. Максимальное значение метрики IOU составило 0.8698. После 60 эпох скорость обучения нейронной сети значительно упала, поэтому тренировка была остановлена.

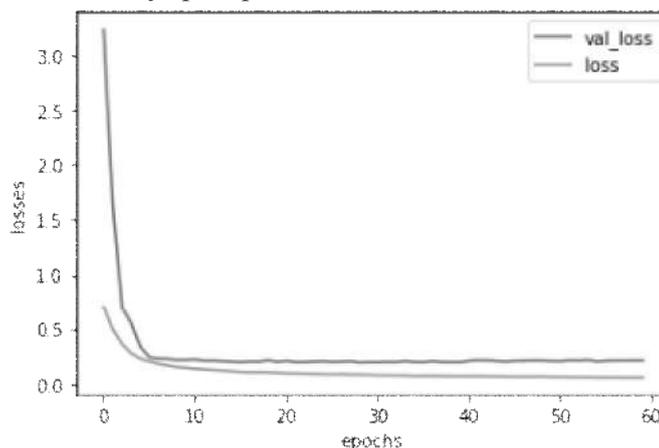


Рис. 3. График ошибки на тренировочной и валидной выборках

5 Обработка видео

Как было указано выше, для тестирования было использовано видео из игры AI DRONE SIMULATOR. Так как нейронная сеть была натренирована на изображении разрешением 256 пикселей на 256 пикселей, для ее качественной работы нужно выбирать картинки со схожим разрешением. Также из-за структуры нейронной сети обработка картинки нейронной сетью со сторонами, не кратными 16, вызывает потерю данных. Поэтому из видеоряда была выделена область 256×192. Для того чтобы определить, безопасна ли посадка в области по дронам размером 50 на 50 сантиметров, проверялось наличие 70% процентов пикселей безопасного класса. На рисунке 4 в первом ряду изображены исходные кадры, во втором — вырезка размером 256×192, в третьем — изображение, сегментированное на два класса.

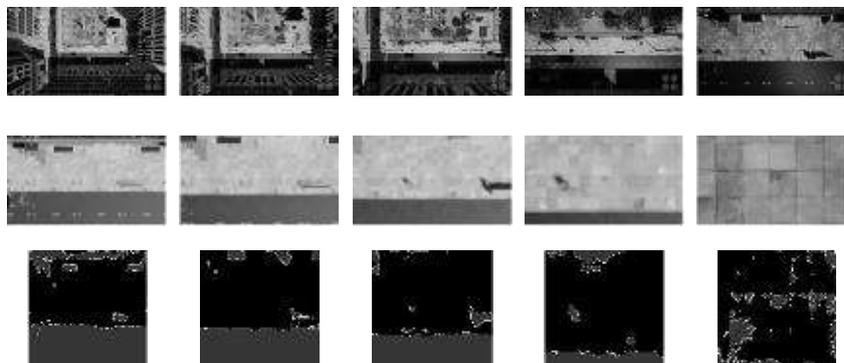


Рис. 4. Кадры полета с тестового видео

Поскольку дрон приземлялся на плитку, место приземления считается полностью безопасным. Приземление происходило в виртуальной среде с идеальными условиями, следовательно, данные могут быть неточными. Однако они позволяют сделать некоторые выводы. На графике точности нейронной сети относительно высоты (рис. 5) отчетливо видно, что на высоте 25–15 метров нейронная сеть показала себя наилучшим образом, поскольку в тестовой выборке изображения были сняты на высоте. Также на высоте 25+ метров заметна слабая отрисовка препятствий в игре, что повлияло на точность сети. Резкий спад точности произошел на высоте 15 метров, что могло быть вызвано отсутствием на этой высоте неблагоприятных для посадки мест. Причина этого в отсутствии в тренировочной выборке изображений с только одним классом. Также это видно на последнем изображении (рис. 4). Исходя из этих выводов процедура посадки должна начинаться на высоте 25 метров от поверхности, окончательно подтверждая посадку на высоте 10–15 метров.

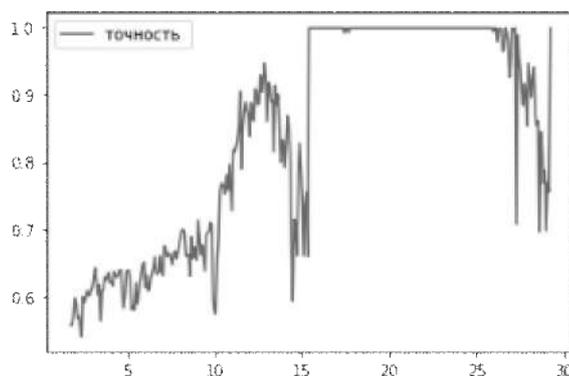


Рис. 5. График точности нейронной сети относительно высоты

Заключение

В статье приведены результаты исследования видов архитектур для сегментации изображения для последующей установки их на беспилотный летательный аппарат. Исследование основывалось на оценке точности и скорости работы разных архитектур для нейронных сетей на валидной выборке тренировочного набора данных, что позволило выбрать наиболее быструю архитектуру с необходимой точностью. Описано применение окружения из игры AI DRONE SIMULATOR. Данные, полученные из виртуального окружения, показали, что использование виртуальных сред для тестирования дронов является возможной альтернативой реальных тестирований. В использованной среде есть ряд недостатков, многие из которых можно исправить, однако для более детальной отладки лучше использовать специально предназначенные для этого среды.

Литература

1. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v5) // UC Berkeley, arXiv:1311.2524v5. URL: <https://arxiv.org/abs/1311.2524> (дата обращения: 10.02.2023).
2. George Vosselman, Gui-Song Xia, Alper Yilmaz, Michael Ying Yang. UAVid: A Semantic Segmentation Dataset for UAV Imagery, arXiv:1810.10438v2. URL: <https://arxiv.org/abs/1810.10438> (дата обращения: 9.02.2023).
3. Shafaei, A., Little, J.J., Schmidt, M.: Play and learn: Using video games to train computer vision models, arXiv:1608.01745v2. URL: <https://arxiv.org/abs/1608.01745> (дата обращения: 11.02.2023).
4. Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for Object Detection, arXiv:1612.03144v2. URL: <https://arxiv.org/abs/1612.03144> (дата обращения: 11.02.2023).
5. Abhishek Chaurasia, Eugenio Culurciello. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation, arXiv:1707.03718v1. URL: <https://arxiv.org/abs/1707.03718>. (дата обращения: 12.02.2023).
6. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597v1. URL: <https://arxiv.org/abs/1505.04597> (дата обращения: 12.02.2023).
7. Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia, Pyramid Scene Parsing Network, arXiv:1612.01105. URL: <https://arxiv.org/abs/1612.01105> (дата обращения: 12.02.2023).
8. Keiron O'Shea, and Ryan Nash. An Introduction to Convolutional Neural Networks, arXiv:1511.08458v2. URL: <https://arxiv.org/abs/1511.08458> (дата обращения: 12.02.2023).
9. Dor Bank, Noam Koenigstein, Raja Giryes. Autoencoders, arXiv:2003.05991. URL: <https://arxiv.org/abs/2003.05991> (дата обращения: 12.02.2023).
10. Sibylle Hess, Wouter Duivesteijn, Decebal Mocanu. Softmax-based Classification is k-means Clustering: Formal Proof, Consequences for Adversarial Attacks, and Improvement through Centroid Based Tailoring, arXiv:2001.01987v1. URL: <https://arxiv.org/abs/2001.01987> (дата обращения: 12.02.2023).
11. Zhilu Zhang, Mert Sabancu. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels, arXiv:1805.07836v4. URL: <https://arxiv.org/abs/1805.07836> (дата обращения: 12.02.2023).

12. Hamid Rezatofighi, Nathan Tsoi JunYoung Gwak Amir Sadeghian, Ian Reid Silvio Savarese. Generalized Intersection over Union, arXiv:1902.09630. URL: <https://arxiv.org/abs/1902.09630>. (дата обращения: 8.02.2023).

13. Marina Sokolova, Nathalie Japkowicz, Stan Szpakowicz. Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation, Advances in Artificial Intelligence. Lecture Notes in Computer Science. 2006. Vol. 4304. P. 1015–1021.

14. Ayoub Benali Amjoud, Mustapha Amrouch, Convolutional Neural Networks Backbones for Object Detection, Image and Signal Processing: 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4–6, 2020, Proceedings, Jun 2020. 282–289.

15. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv:1704.04861. <https://arxiv.org/abs/1704.04861> (дата обращения: 12.02.2023).

16. Обнаружение лесных пожаров с помощью беспилотных летательных аппаратов / О. В. Кузьмин, А. А. Лавлинская, Г. А. Филь, М. Д. Камнев // Математика, ее приложения и математическое образование (МПМО'20): материалы конференции. 2020. С. 134–136.

17. Лавлинская А. А., Филь Г. А., Камнев М. Д. Создание модели квадрокоптера-эколога // Прикладные вопросы дискретного анализа: сб. науч. тр. / под редакцией О. В. Кузьмина. Иркутск: Изд-во ИГУ, 2020. Вып. 6. С. 78–83.

18. Проектирование БПЛА для мониторинга лесных массивов и водных пространств / О. В. Кузьмин, А. А. Лавлинская, Б. А. Тараканов [и др.] // Прикладные проблемы дискретного анализа: сб. науч. тр. / под редакцией О. В. Кузьмина. Иркутск: Изд-во ИГУ, 2021. Вып. 7. С. 66–72.

19. Кузьмин О. В., Лавлинский М. В. Создание модели беспилотного летательного аппарата для помощи в решении проблемы пожаров в Иркутской области // Современные технологии. Системный анализ. Моделирование. 2020. № 2 (66). С. 136–143. DOI: 10.26731/1813-9108.2020.2(66).136-143.

20. Кузьмин О. В., Лавлинский М. В. Создание модели беспилотного летательного аппарата типа биплан, использующей генетический алгоритм в полете, для помощи в решении проблемы пожаров в Иркутской области // Современные технологии. Системный анализ. Моделирование. 2021. № 3 (71). С. 185–192. DOI: 10.26731/1813-9108.2021.3(71).185-192.

Статья поступила в редакцию 08.02.2023; одобрена после рецензирования 10.03.2023; принята к публикации 13.03.2023.

IMAGE SEGMENTATION ARCHITECTURES AND THEIR PRACTICAL APPLICATION FOR UNMANNED AIRCRAFT LANDING

Oleg V. Kuzmin

Dr. Sci. (Phys. and Math.), Professor,
Head of the Department of Probability Theory and Discrete Mathematics
Irkutsk State University
1 Karl Marx St., Irkutsk 664003, Russia

Maxim V. Lavlinsky
graduate student,
Irkutsk State University
1 Karl Marx St., Irkutsk 664003, Russia

Abstract: This paper discusses image segmentation architectures and their practical application for unmanned landing, examples of implemented systems for delivering goods by drones. The optimal drone landing was selected and implemented using image segmentation by a neural network. The data set for training is selected and modified by mirroring the image and rotating the image. An additional data sample was created using an application that simulates the flight of a drone in a custom environment. Found a way to quickly mark up a video recording of a drone landing.

Neural networks with different architectures were tested to find the best one for needed purpose. For this task four trainings of different models were done to find needed accuracy and speed of analyzing one frame of a video. Two different metrics of accuracy were used for neural network to get reliable data of recognition. The best environment classes of recognition were selected. Three models with different sets of classes were created. Results were analyzed and the best classes were selected.

The final model of the neural network was done on the basis of linknet architecture. For best performance additional pretrained weights were chosen. Training of neural network had been going for 13 hours

Video of a drone landing was processed in a virtual environment. Based on height of a drone landing spot and area of following segmentation was chosen. The optimal height above the surface to start the landing process. Was made a demonstrative video of drone landing in optimal conditions.

Keywords: neural network, image segmentation, unmanned drone, unmanned landing, remote control.

For citation

Kuzmin O. V., Lavlinsky M. V. Image Segmentation Architectures and Their Practical Application for Unmanned Aircraft Landing // Bulletin of Buryat State University. Mathematics, Informatics. 2023. N. 1. P. 37–46.

The article was submitted 08.02.2023; approved after reviewing 10.03.2023; accepted for publication 13.03.2023.