УДК 004.931
doi:10.21685/2227-8486-2021-3-9

МЕТОД ПРОГНОЗИРОВАНИЯ КАДРОВ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ НА ОСНОВЕ ГЕНЕРАТИВНЫХ НЕЙРОННЫХ СЕТЕЙ

В. Д. Сазыкина[1], М. А. Митрохин[2]

[1, 2] Пензенский государственный университет, Пенза, Россия
[1] sazykinavd@yandex.ru, [2] mmax83@mail.ru

Аннотация. *Актуальность и цели*. Рассматриваются основные недостатки традиционных подходов к обнаружению движущихся объектов в видеопотоке. Обосновывается необходимость использования новых подходов на базе глубокого обучения. *Материалы и методы*. В качестве перспективного направления в решении задачи обнаружения движущихся объектов в видеопотоке предлагается использовать генеративную состязательную сеть. Для сохранения семантически в процессе нормализации – метод пространственно-адаптивной нормализации. Совместно с методом пространственно-адаптивной нормализации предлагается применить метод семантической сегментации и метод оценки оптического потока. *Результаты*. В результате исследования был разработан метод прогнозирования кадров. Предложено использовать блоки Multi-SPADE и повторное применение сети деформационных объемов «Devon» к спрогнозированному и реальному, смежному во времени кадрам. *Выводы*. Предложенный метод прогнозирования кадров видеопоследовательности может служить основой для построения метода обнаружения движущихся объектов.

Ключевые слова: глубокое обучение, генеративная состязательная сеть, пространственно-адаптивная нормализация, карта пространства, семантическая сегментация, сеть встраивания меток, оптический поток, сеть деформационных объемов

Для цитирования: Сазыкина В. Д., Митрохин М. А. Метод прогнозирования кадров видеопоследовательности на основе генеративных нейронных сетей // Модели, системы, сети в экономике, технике, природе и обществе. 2021. № 3. С. 91–97. doi:10.21685/2227-8486-2021-3-9

METHOD OF PREDICTION OF VIDEO SEQUENCE FRAMES BASED ON GENERATIVE NEURAL NETWORKS

V.D. Sazykina[1], M.A. Mitrokhin[2]

[1, 2] Penza State University, Penza, Russia
[1] sazykinavd@yandex.ru, [2] mmax83@mail.ru

Abstract. *Background*. The main disadvantages of traditional approaches to detecting moving objects in a video stream are considered. The need for new approaches based on in-depth learning is justified. *Materials and methods*. As a promising direction in solving the problem of detecting moving objects in a video stream, the use of a generative adversarial network is proposed. To preserve semantically in the process of normalization a method

of spatial-adaptive normalization is proposed. Together with the method of spatial-adaptive normalization, it is proposed to use the method of semantic segmentation and the method of estimating optical flow. *Results*. As a result of the research, a method of forecasting video frames was developed. It is proposed to use Multi-SPADE blocks, and the repeated application of the "Devon" network of deformation volumes to the predicted frame and the real one, adjacent in time. *Conclusions*. The proposed method for predicting frames of a video sequence can be used for constructing a method for detecting moving objects.

Keywords: deep learning, generative adversarial network, spatial-adaptive normalization, space map, semantic segmentation, label embedding network, optical flow, deformable volume network

For citation: Sazykina V.D., Mitrokhin M.A. Method of prediction of video sequence frames based on generative neural networks. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve = Models, systems, networks in economics, technology, nature and society*. 2021;(3): 91–97. (In Russ.). doi:10.21685/2227-8486-2021-3-9

Введение

Традиционные подходы к обнаружению движущихся объектов в видеопотоке обычно сводятся к сравнению пикселей последовательности кадров [1] и вычислению оптического потока [2]. Повысить эффективность обнаружения объектов позволяют подходы, основанные на предсказании движения пикселей в будущем кадре, предполагающие известным закон движения [3]. Одним из основных недостатков этих подходов является то, что прогнозирование движения подвержено ошибкам из-за различных факторов, таких как освещенность, нелинейность движения объектов, сложный фон и т.д. Существует ряд методов, основанных на глубоком обучении, наследующих эту идею. Глубокие сети показывают удовлетворительные результаты при обработке сложных движений в динамической сцене.

Прогнозирование не отдельных частей, а целиком кадра видеопоследовательности – это многообещающее направление исследований в области компьютерного зрения. Прогнозирование будущих кадров из набора последовательно предшествующих может использоваться для обнаружения движущихся объектов и аномальных событий. Данная задача имеет широкую прикладную ценность в таких областях, как робототехника и автономное вождение.

Материалы и методы

Основным этапом решения задачи прогнозирования кадра является выбор архитектуры нейросети.

Генеративная состязательная сеть (GAN) – это среда машинного обучения для генерации данных, обучающаяся генерировать новые данные, которые следуют распределению, аналогичному распределению в обучающих наборах. GAN включает в себя две состязательные сети, сеть генератора и сеть дискриминатора. Генератор стремится сгенерировать правдоподобный выходной сигнал, чтобы «обмануть» дискриминатор, в то время как дискриминатор пытается правильно классифицировать сгенерированные данные [4]. Классическая схема GAN-архитектуры представлена на рис. 1.

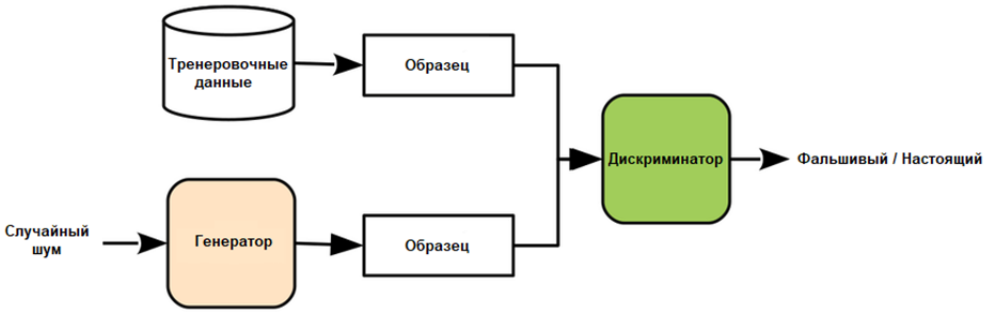

Рис. 1. Схема архитектуры GAN

Одним из самых популярных типов GAN, в частности для генерации изображений, стала архитектура глубокой сверточной генеративной состязательной сети (DCGAN) [5]. Именно данную архитектуру предлагается использовать для дальнейшего исследования. Ключевым отличием архитектуры DCGAN от GAN является применение модели сети для генератора и дискриминатора – глубоких сверточных сетей. Схемы генератора и дискриминатора DCGAN представлены на рис. 2 и 3 соответственно.

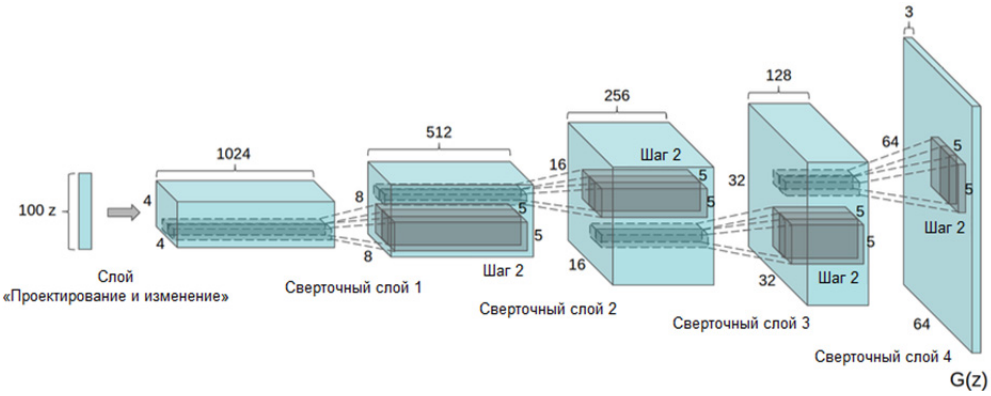

Рис. 2. Схема генератора архитектуры DCGAN

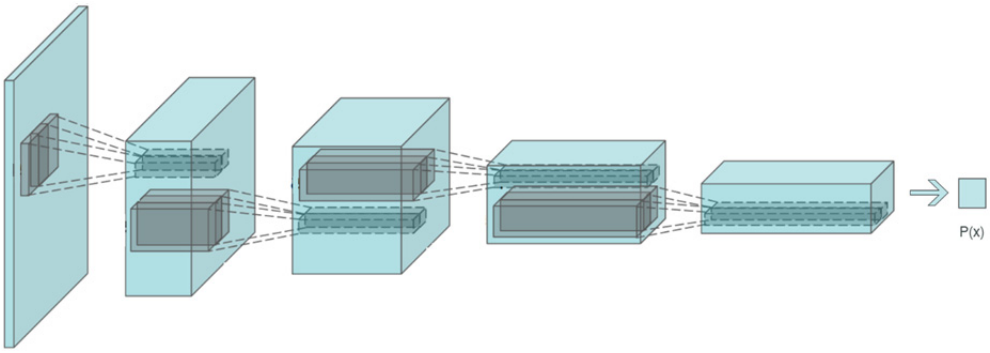

Рис. 3. Схема дискриминатора архитектуры DCGAN

В архитектуре DCGAN семантический макет напрямую передается в качестве входных данных в глубокую сеть, которые затем обрабатываются с помощью стеков слоев свертки, нормализации и нелинейности. Согласно

исследованиям [6] это не оптимально, поскольку слои нормализации имеют тенденцию «смывать» семантическую информацию. Чтобы решить эту проблему, в работе [6] предлагается использовать метод пространственно-адаптивной нормализации (SPADE), схема которого представлена на рис. 4.

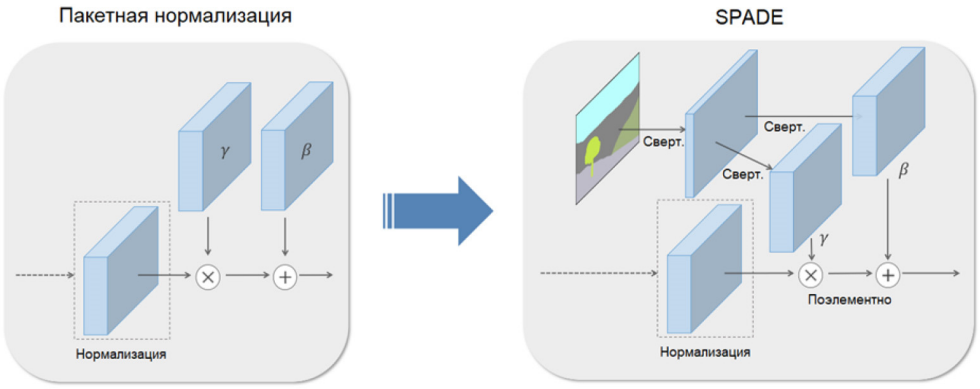

Рис. 4. Схема метода SPADE

Данный метод предполагает использование макета ввода для модуляции активаций в слоях нормализации с помощью пространственно-адаптивного, обученного преобразования. Согласно исследованию [6] применение пространственно-адаптивной нормализации позволяет синтезировать значительно лучшие результаты по сравнению с другими методами нормализации.

В качестве основного компонента генератора возможно использование блоков Multi-SPADE, каждый из которых состоит из нескольких SPADE-слоев. Предполагается, что каждый SPADE-слой принимает на вход карту пространства: семантическую сегментацию, оптическую деформацию и выход предыдущего слоя в виде оптического потока. Затем внутри слоя применяются соответствующие преобразования на промежуточных картах признаков. Схема применения Multi-SPADE блоков представлена на рис. 5.

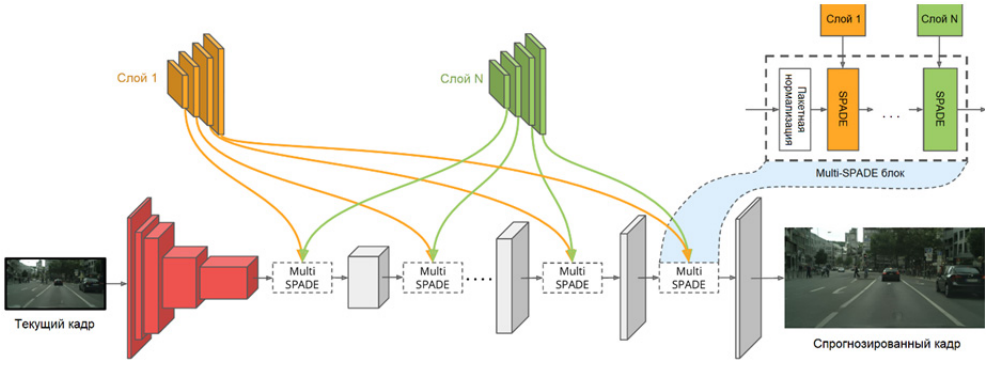

Рис. 5. Схема применения Multi-SPADE блоков

В качестве подхода семантической сегментации предлагается применить метод под названием «Label Embedding Network» [7], который может изучать представление меток (встраивание меток) в процессе обучения глубоких сетей. С помощью предлагаемого метода встраивание меток адаптивно и автоматически осуществляется посредством обратного распространения оши-

бок. Функция потерь, представленная исходным унитарным кодом, преобразуется в новую функцию потерь с «мягким» распределением, так что изначально несвязанные метки постоянно взаимодействуют друг с другом в процессе обучения. В результате обученная модель может достичь существенно более высокой точности и более высокой скорости сходимости [7]. Согласно исследованию [7] экспериментальные результаты, основанные на соревновательных задачах, демонстрируют эффективность применения данного метода.

Современные модели нейронных сетей оценивают оптический поток большого смещения в нескольких разрешениях и используют деформацию для распространения оценки между двумя разрешениями. Несмотря на их впечатляющие результаты, известно, что у этого подхода есть две проблемы. Во-первых, оценка оптического потока с несколькими разрешениями не работает в ситуациях, когда небольшие объекты движутся быстро. Во-вторых, деформация создает артефакты, когда происходит окклюзия или дисклюзия. Для решения данных проблем в качестве подхода определения деформации предлагается применить сеть деформационных объемов «Deformable Volume Network – Devon» [8], которая может оценивать разномасштабный оптический поток в едином высоком разрешении на эффективном уровне [8].

Результаты

В результате применения и объединения перечисленных подходов в контексте обнаружения движущегося объекта разработан метод прогнозирования кадра видеопоследовательности. Общая схема прогнозирования кадра видеопоследовательности с использованием всех рассмотренных подходов представлена на рис. 6.

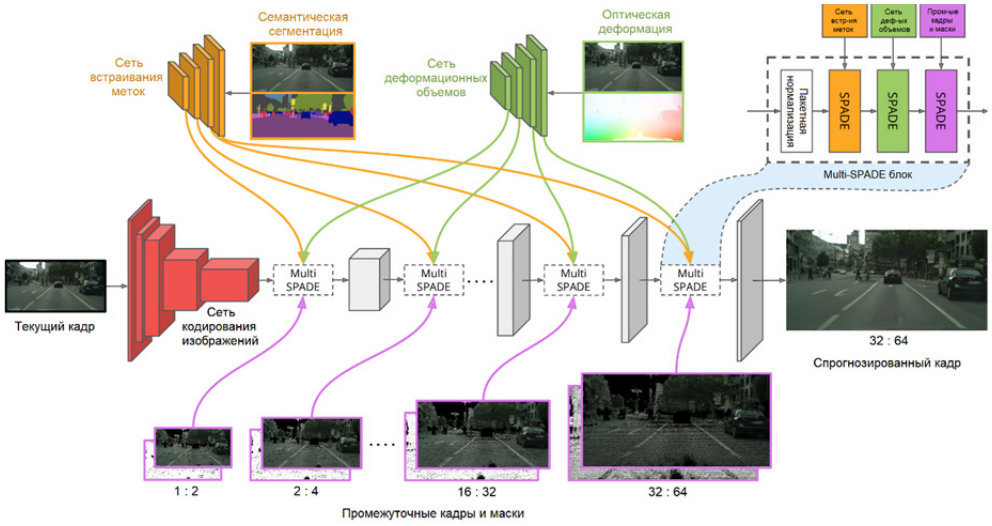

Рис. 6. Общая схема прогнозирования кадра видеопоследовательности с использованием нейронных сетей

Метод базируется на использовании блоков Multi-SPADE, слои которых последовательно принимают на вход карту пространства: семантическую сегментацию, оптическую деформацию и выход предыдущего слоя в виде оптического потока. Ключевым подходом для обнаружения движущегося объекта в видеопотоке является повторное применение сети деформацион-

ных объемов «Devon» к спрогнозированному и реальному, смежному во времени кадрам.

Реализация разработанного метода возможна с использованием таких инструментов, как кроссплатформенный фреймворк Qt, библиотека машинного обучения libtorch с поддержкой Nvidia CUDA [9, 10], а также библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения – OpenCV [1]. Дальнейшее исследование направлено на определение эффективности предлагаемого метода и конечной реализации.

Заключение

Рассмотрено применение глубокого обучения, в частности архитектуры DCGAN, в качестве подхода к обнаружению движущихся объектов в видеопотоке. Для основных компонентов архитектуры выполнен анализ современных подходов, выбраны методы и решения, позволяющие повысить эффективность обнаружения объектов. Предполагаемая эффективность базируется на результатах исследований, посвященных областям, соответствующим методам. В результате применения выбранных подходов разработан метод обнаружения движущихся объектов в видеопотоке. В качестве вектора дальнейшего исследования предложены реализация разработанного метода и получение данных, позволяющих выполнить сравнительный анализ с существующими аналогами.

Список литературы

1. Гарсия Б. Г., Суарес Д. О., Аранда Э. Л. [и др.]. Обработка изображений с помощью OpenCV. М. : ДМК, 2016. 210 с.
2. Fleet D. J., Weiss Y. Optical Flow Estimation // Handbook of Mathematical Models in Computer Vision, 2006. URL: https://www.cs.toronto.edu/~fleet/research/Papers/flowChapter05.pdf (дата обращения: 08.07.2021).
3. Сазыкина В. Д., Митрохин М. А. Обработка изображений с динамическим фоном // Новые информационные технологии и системы : сб. науч. ст. XVI Междунар. науч.-техн. конф. (г. Пенза, 27–29 ноября 2019 г.). Пенза : Изд-во ПГУ, 2019. 312 с.
4. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B. Generative Adversarial Networks, 2014. URL: https://arxiv.org/abs/1406.2661 (дата обращения: 08.07.2021).
5. Radford A., Metz L., Chintala S. Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks, 2015. URL: https://arxiv.org/abs/1511.06434 (дата обращения: 08.07.2021).
6. Park T., Liu M., Wang T., Zhu J. Semantic Image Synthesis with Spatially-Adaptive Normalization (SPADE), 2019. URL: https://arxiv.org/abs/1903.07291 (дата обращения: 08.07.2021).
7. Sun X., Wei B., Ren X., Ma S. Label Embedding Network: Label Embedding Network: Learning Label Representation For Soft Training Of Deep Networks, 2017. URL: https://arxiv.org/abs/1710.10393 (дата обращения: 08.07.2021).
8. Lu Y., Valmadre J., Wang H. [et all.]. Devon: Deformable Volume Network for Learning Optical Flow, 2020. URL: https://arxiv.org/abs/1802.07351 (дата обращения: 08.07.2021).
9. Макмахан Б., Рао Д. Знакомство с PyTorch. СПб. : Питер, 2020. 256 с.
10. PyTorch C++ API. URL: https://pytorch.org/cppdocs/ (дата обращения: 08.07.2021).

References

1. Garsiya B.G., Suares D.O., Aranda E.L. [et al.]. *Obrabotka izobrazheniy s pomoshch'yu OpenCV = Image processing with OpenCV*. Moscow: DMK, 2016:210. (In Russ.)
2. Fleet D.J., Weiss Y. Optical Flow Estimation. *Handbook of Mathematical Models in Computer Vision, 2006*. Available at: https://www.cs.toronto.edu/~fleet/research/Papers/flowChapter05.pdf (accessed 08.07.2021).
3. Sazykina V.D., Mitrokhin M.A. Image processing with dynamic background. *Novye informatsionnye tekhnologii i sistemy: sb. nauch. st. XVI Mezhdunar. nauch.-tekhn. konf. (g. Penza, 27–29 noyabrya 2019 g.) = New information technologies and systems : collection of scientific Articles XVI Inter-dunar. scientific-technical. conf.* Penza: Izd-vo PGU, 2019:312. (In Russ.)
4. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B. *Generative Adversarial Networks, 2014*. Available at: https://arxiv.org/abs/1406.2661 (accessed 08.07.2021).
5. Radford A., Metz L., Chintala S. *Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks, 2015*. Available at: https://arxiv.org/abs/ 1511.06434 (accessed 08.07.2021).
6. Park T., Liu M., Wang T., Zhu J. *Semantic Image Synthesis with Spatially-Adaptive Normalization (SPADE), 2019*. Available at: https://arxiv.org/abs/1903.07291 (accessed 08.07.2021).
7. Sun X., Wei B., Ren X., Ma S. *Label Embedding Network: Label Embedding Network: Learning Label Representation For Soft Training Of Deep Networks, 2017*. Available at: https://arxiv.org/abs/1710.10393 (accessed 08.07.2021).
8. Lu Y., Valmadre J., Wang H. [et al.]. *Devon: Deformable Volume Network for Learning Optical Flow, 2020*. Available at: https://arxiv.org/abs/1802.07351 (accessed 08.07.2021).
9. Makmakhan B., Rao D. *Znakomstvo s PyTorch = Acquaintance with PyTorch*. Saint Petersburg: Piter, 2020:256. (In Russ.)
10. *PyTorch C++ API*. Available at: https://pytorch.org/cppdocs/ (accessed 08.07.2021).

Информация об авторах / Information about the authors

Валерия Дмитриевна Сазыкина
аспирант,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: sazykinavd@yandex.ru

Максим Александрович Митрохин
доктор технических наук, профессор,
заведующий кафедрой
вычислительной техники,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: mmax83@mail.ru

Valeriya D. Sazykina
Postgraduate student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Maxim A. Mitrokhin
Doctor of technical sciences, professor,
head of the sub-department
of computer engineering,
Penza State University
(40 Krasnaya street, Penza, Russia)

Авторы заявляют об отсутствии конфликта интересов /
The authors declare no conflicts of interests.

Поступила в редакцию/Received 30.04.2021
Поступила после рецензирования/Revised 26.06.2021
Принята к публикации/Accepted 12.07.2021