

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СОВРЕМЕННЫХ АРХИТЕКТУР НЕЙРОННЫХ СЕТЕЙ ДЛЯ ЗАДАЧ КОМПЬЮТЕРНОГО ЗРЕНИЯ

COMPARATIVE ANALYSIS OF MODERN NEURAL NETWORK ARCHITECTURES FOR COMPUTER VISION TASKS

R. Nazipov

Summary. This article presents a comparative analysis of three modern neural network architectures for computer vision tasks: Convolutional Neural Networks (CNNs), ResNet, and YOLO. The key features, advantages, and limitations of each architecture are examined. The practical significance and future development prospects of neural networks in the field of computer vision are discussed, including the development of hybrid models, the use of transfer learning methods, and integration with classical approaches. The importance of further research to improve the efficiency, adaptability, and interpretability of neural networks in solving a wide range of computer vision problems is emphasized.

Keywords: computer vision, neural networks, convolutional neural networks, ResNet, YOLO, neural network architectures, deep learning, interpretability, transfer learning, hybrid models.

Назипов Рустам Салаватович
руководитель НИИ «ЭВРИКА», г. Казань
rustam.nazipov@gmail.com

Аннотация. В данной статье проводится сравнительный анализ трех современных архитектур нейронных сетей для задач компьютерного зрения: сверточных нейронных сетей (CNN), ResNet и YOLO. Рассматриваются ключевые особенности, преимущества и ограничения каждой архитектуры. Обсуждаются практическая значимость и перспективы развития нейронных сетей в области компьютерного зрения, включая разработку гибридных моделей, использование методов переноса знаний и интеграцию с классическими подходами. Подчеркивается важность дальнейших исследований для повышения эффективности, адаптивности и интерпретируемости нейронных сетей в решении широкого спектра задач компьютерного зрения.

Ключевые слова: компьютерное зрение, нейронные сети, сверточные нейронные сети, ResNet, YOLO, архитектуры нейронных сетей, глубокое обучение, интерпретируемость, перенос знаний, гибридные модели.

Введение

Компьютерное зрение — одна из наиболее динамично развивающихся областей искусственного интеллекта, находящая применение в различных сферах деятельности, таких как медицина, промышленность, робототехника и безопасность [1]. Значительный прогресс в этой области был достигнут благодаря развитию нейронных сетей, способных эффективно решать задачи классификации, детектирования и сегментации изображений [2]. Выбор оптимальной архитектуры нейронной сети для конкретной задачи компьютерного зрения является важным фактором, определяющим качество и скорость работы системы [3].

В данной статье проводится сравнительный анализ трех современных архитектур нейронных сетей: сверточных нейронных сетей (CNN), ResNet и YOLO. Цель анализа — выявить их ключевые особенности, преимущества и ограничения, а также определить наиболее перспективные направления дальнейшего развития. Понимание этих аспектов позволит исследователям и практикам в области компьютерного зрения выбирать наиболее подходящие архитектуры для решения своих задач и разрабатывать новые, более эффективные модели.

Сверточные нейронные сети (CNN)

Сверточные нейронные сети (CNN) — это специализированный тип нейронных сетей, разработанный для

эффективной обработки изображений [4]. Ключевой особенностью CNN является наличие сверточных слоев, которые позволяют извлекать иерархические признаки из изображений с помощью операции свертки [5]. Эта операция обеспечивает инвариантность к небольшим сдвигам и искажениям входных данных, что делает CNN устойчивыми к изменениям положения и ориентации объектов на изображениях [6].

Преимущества CNN:

- Способность автоматически обучаться иерархическим признакам без необходимости в ручном проектировании [5].
- Высокая точность в задачах классификации изображений [7].
- Устойчивость к небольшим сдвигам и искажениям входных данных [6].

Недостатки CNN:

- Требуют большого объема обучающих данных и вычислительных ресурсов [8].
- Могут испытывать трудности с обобщением на новые данные [9].

Несмотря на эти ограничения, CNN остаются одной из наиболее популярных и эффективных архитектур для решения задач компьютерного зрения, особенно в области классификации изображений.

Архитектура ResNet

ResNet (Residual Network) — это глубокая архитектура нейронной сети, предложенная для решения проблемы исчезающих градиентов при обучении очень глубоких сетей [2]. Ключевой особенностью ResNet является использование остаточных блоков, которые позволяют эффективно обучать сети с сотнями и даже тысячами слоев. Эти блоки содержат прямые соединения (shortcut connections), которые пропускают один или несколько слоев и суммируют их выход с выходом предыдущих слоев. Такая структура позволяет градиентам беспрепятственно распространяться по сети, облегчая процесс обучения [2].

Преимущества ResNet:

- Возможность обучения очень глубоких моделей без потери производительности [2].
- Высокая точность на различных задачах компьютерного зрения [10].
- Хорошая переносимость предобученных моделей на новые задачи (трансферное обучение) [9].

Недостатки ResNet:

- Высокие требования к вычислительным ресурсам и времени обучения [8].
- Сложность интерпретации полученных моделей из-за их глубины [3].

Архитектура ResNet получила широкое признание в сообществе компьютерного зрения и стала основой для многих современных моделей, достигающих впечатляющих результатов на различных задачах.

Архитектура YOLO

YOLO (You Only Look Once) — это архитектура нейронной сети, разработанная специально для задачи детектирования объектов в реальном времени [4]. В отличие от традиционных подходов, YOLO обрабатывает изображение целиком и напрямую предсказывает ограничивающие рамки и классы объектов. Это позволяет достичь высокой скорости работы, сохраняя при этом достаточный уровень точности.

Ключевой особенностью YOLO является разделение входного изображения на сетку и предсказание ограничивающих рамок и вероятностей классов для каждой ячейки сетки. Затем эти предсказания комбинируются для получения финального результата детектирования [4]. Такой подход позволяет YOLO эффективно обрабатывать изображения различных размеров и обнаруживать объекты разных масштабов и соотношений сторон.

Преимущества YOLO:

- Высокая скорость работы, подходящая для приложений реального времени [4].

- Простота архитектуры и обучения по сравнению с другими методами детектирования [4].
- Способность обнаруживать объекты различных масштабов и соотношений сторон [6].

Недостатки YOLO:

- Более низкая точность по сравнению с некоторыми двухэтапными подходами [7].
- Трудности с обнаружением малых и близко расположенных объектов [10].

Несмотря на эти ограничения, YOLO остается одной из наиболее популярных архитектур для задач детектирования объектов в реальном времени благодаря своей скорости и простоте.

Практическая значимость и перспективы развития

Проведенный анализ показывает, что каждая из рассмотренных архитектур нейронных сетей имеет свои сильные стороны и области применения. Выбор оптимальной архитектуры зависит от конкретных требований задачи, таких как точность, скорость работы и доступные вычислительные ресурсы. В практических приложениях компьютерного зрения, требующих высокой точности классификации изображений, целесообразно использовать CNN или ResNet. Для задач детектирования объектов в реальном времени, например, в системах видеонаблюдения или беспилотных автомобилях, предпочтительным выбором может стать YOLO.

Перспективы развития архитектур нейронных сетей для компьютерного зрения связаны с разработкой более эффективных моделей, способных обучаться на ограниченных объемах данных и обеспечивать высокую точность при меньших вычислительных затратах. Одним из многообещающих направлений является использование методов переноса знаний и мета-обучения, которые позволяют адаптировать предобученные модели к новым задачам с минимальным объемом дополнительных данных [9]. Кроме того, важным направлением развития является повышение интерпретируемости и объяснимости нейронных сетей, что позволит лучше понимать принципы их работы и повысит доверие к полученным результатам [3].

Также перспективным направлением является разработка гибридных архитектур, объединяющих преимущества различных подходов. Например, сочетание сверточных слоев CNN с механизмами внимания из трансформеров может позволить создавать более мощные и гибкие модели для различных задач компьютерного зрения [1]. Кроме того, интеграция нейронных сетей с классическими методами компьютерного зрения и использование физических и геометрических ограни-

чений может повысить надежность и интерпретируемость результатов [5].

Заключение

В данной статье был проведен сравнительный анализ трех современных архитектур нейронных сетей для задач компьютерного зрения: сверточных нейронных сетей (CNN), ResNet и YOLO. Каждая из этих архитектур имеет свои преимущества и недостатки, которые необходимо учитывать при выборе оптимальной модели для конкретной задачи. CNN хорошо подходят для задач классификации изображений, ResNet позволяют обучать очень глубокие модели с высокой точностью, а YOLO обеспечивает высокую скорость детектирования объектов в реальном времени.

Дальнейшее развитие архитектур нейронных сетей будет направлено на повышение их эффективности, адаптивности и интерпретируемости. Разработка ги-

бридных моделей, использование методов переноса знаний и интеграция с классическими подходами компьютерного зрения позволят создавать более мощные, надежные и объяснимые системы. Это расширит возможности применения нейронных сетей в различных прикладных областях компьютерного зрения и ускорит внедрение технологий искусственного интеллекта в повседневную жизнь.

Несомненно, компьютерное зрение и нейронные сети будут играть все более важную роль в развитии технологий будущего. Понимание особенностей современных архитектур и перспектив их развития является ключом к эффективному использованию этих мощных инструментов для решения широкого спектра задач — от медицинской диагностики до автономного вождения. Дальнейшие исследования в этой области помогут раскрыть весь потенциал нейронных сетей и компьютерного зрения на благо общества.

ЛИТЕРАТУРА

1. Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018.
2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
3. Zhang, Q.S., & Zhu, S.C. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 27–39.
4. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).
5. Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8), 5455–5516.
6. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).
7. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... & Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7310–7311).
8. Canziani, A., Paszke, A., & Culurciello, E. (2016). An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678*.
9. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. *Advances in neural information processing systems*, 27.
10. Bochkovskiy, A., Wang, C.Y., & Liao, H.Y.M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.