УДК 004.032.26

ПРОБЛЕМЫ «УМИРАЮЩИХ НЕЙРОНОВ» И «ЗАТУХАЮЩИХ ГРАДИЕНТОВ» В СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЯХ

PROBLEMS OF «DYING NEURONS» AND «FADING GRADIENTS» IN CONVOLUTIONAL NEURAL NETWORKS

Лесовой Эдуард Васильевич
аспирант, инженер кафедры автоматики, телемеханики, связь и вычислительная техника,
Донецкий институт железнодорожного транспорта
ed.lesovoi2014@yandex.ru

Аннотация. Современные глубокие нейронные сети активно применяются в задачах компьютерного зрения, где устойчивость обучения напрямую зависит от выбора функции активации. В работе выполнен сравнительный анализ функций активации ReLU, Sigmoid, ELU, SeLU и PeLU с позиции распространения градиента и предотвращения деградации активности нейронов. Рассмотрены аналитические выражения функций и их производных, исследованы особенности поведения в положительной и отрицательной областях аргумента. Показано влияние функций активации на стабильность обучения, вычислительную эффективность и вероятность возникновения эффектов затухающего градиента и умирающих нейронов.

Ключевые слова: функции активации; глубокие нейронные сети; затухающий градиент; умирающие нейроны; ReLU; ELU; Sigmoid, SeLU, PeLU.

Lesovoy Eduard Vasilyevich
Post-graduate student,
Engineer of the Department of Automation, Telemechanics, Communications and Computing,
Donetsk Institute of Railway Transport
ed.lesovoi2014@yandex.ru

Annotation. Modern deep neural networks are actively used in computer vision tasks, where the stability of learning directly depends on the choice of the activation function. The paper provides a comparative analysis of the activation functions of ReLU, Sigmoid, ELU, SeLU and PeLU from the perspective of gradient propagation and prevention of degradation of neuronal activity. The analytical expressions of functions and their derivatives are considered, the features of behavior in the positive and negative areas of the argument are investigated. The effect of activation functions on learning stability, computational efficiency, and the likelihood of effects of a decaying gradient and dying neurons is shown.

Keywords: activation functions; deep neural networks; fading gradient; dying neurons; ReLU; ELU; Sigmoid, SeLU, PeLU.

Как известно в последнее время во всех сферах жизни человека все чаще применяют искусственная нейронная сеть (ИНС). Не исключениям является, и транспортная отрасль, в которой активно внедряются инструменты машинного зрения. [1–3]

Интенсивное внедрение искусственных нейронных сетей в задачи наблюдения, обработки изображений и автоматического управления [4–6] ведёт к возрастанию требований к надёжности обучения глубоких нейронных сетей.

Особенно важным это становится для сверточных нейронных сетей, которые широко применяются в системах машинного зрения в транспортной отрасли.

Ключевым компонентом любой нейронной сети считается функция активации [7], обеспечивающая нелинейность модели и позволяющая приближать сложные взаимосвязи между входными и выходными данными.

Тем не менее, с увеличением глубины сети возникают серьёзные проблемы обучения – эффекты затухания градиента и «умирающих» нейронов, которые значительно осложняют процесс оптимизации.

Самая популярной функцией активации ReLU, благодаря своей простоте вычислений и высокой скорости работы [8].

Однако её использование связано с потерей активности у части нейронов. Другие функции активации устраняют этот недостаток, но могут приводить к ухудшению распространения градиента в сети.

Целью работы является сравнительный анализ наиболее популярных свойств функций активации, которые применяются в скрытых слоях: ReLU, Sigmoid, ELU, SeLU и PeLU в сверхточных нейронных сетях с точки зрения их влияния на распространение градиента и устойчивость обучения глубоких нейронных сетей.

Для достижения поставленной цели формулируются следующие задачи:

– проанализировать аналитические выражения функций и их производных;
– исследовать поведение активаций в отрицательной и положительной областях аргумента;
– выдвинуть гипотезу о преимуществах выбранные функции активации в предотвращении деградации активности;
– провести валедацию на специальном программном обеспечении для проверки сформулированных предположений.

Функции активации играют ключевую роль в работе искусственных нейронных сетей, обеспечивая введение нелинейности и возможность моделирования сложных зависимостей в данных. Их выбор напрямую влияет на стабильность обучения, скорость сходимости и распространение градиента.

Функция выпрямленной линейной единицы (ReLU) широко применяется в скрытых слоях глубоких нейронных сетей благодаря вычислительной простоте и эффективности обучения. Она определяется выражением:

$$f'_{\mathrm{Re}LU}(x) = \max(0, x), \tag{1}$$

где x – вход нейрона ($x = w \times z + b$); w – вектор весовых коэффициентов; z – входной вектор значений; b – смещение.

ReLU обеспечивает линейный рост при положительных значениях аргумента и нулевой выход при отрицательных. Это ускоряет вычисления и снижает вероятность возникновения проблемы затухающего градиента.

Однако при отрицательных входах производная функции равна нулю, что приводит к эффекту «умирающих нейронов», когда веса перестают обновляться в процессе обратного распространения ошибки.

Альтернативным решением данной проблемы является область определения за счет неплотности и не линейности, как это реализовано в функции ELU. Экспоненциальная линейная функция ELU (Exponential Linear Unit) была предложена для уменьшения недостатков ReLU. Она задаётся следующим образом:

$$f'_{ELU}(x) = \begin{cases} x, & x > 0 \\ \alpha\left(e^{x} - 1\right), & x \le 0 \end{cases}, \tag{2}$$

где α – гиперпараметр, задающий глубину насыщения в отрицательной области ($\alpha > 0$); e – основание натурального логарифма (e = 2.718).

В отличие от ReLU, ELU сохраняет ненулевую производную в отрицательной области, что позволяет продолжать корректировку весов. Дополнительным преимуществом является возможность получения отрицательных выходных значений, благодаря чему средние активации приближаются к нулю и обучение ускоряется.

Недостатками являются более высокая вычислительная сложность и зависимость от выбора параметра α. ELU лучше справляется с проблемой умирающих нейронов в отрицательных значениях, но ее проблема самонормализации. Смонормализация - по мере прохождения данных по сети активации, как правило, остаются близкими к среднему значению, равному нулю, и стандартному отклонению, равному единице, без дополнительных методов нормализации. Решением этой проблемы является функция активации SeLU.

SeLU представляет масштабированную модификацию ELU и обеспечивает свойство самонормализации:

$$f'_{SeLU}(x) = \lambda \begin{cases} x, & x > 0 \\ \alpha\left(e^{x} - 1\right), & x \le 0 \end{cases}, \tag{3}$$

где λ – коэффициент масштабирования ($\lambda \approx 1.0507$); α – параметр экспоненциальной части ($\alpha \approx 1.6733$); x – вход нейрона.

При определённых условиях распределение активаций автоматически стабилизируется, уменьшая внутренний сдвиг данных и повышая устойчивость обучения глубоких сетей.

Ограничением является фиксированность коэффициентов масштабирования.

Стоит заметить во многих задачах обучения имеется потребность регулирование коэффициент наклона. В этом случае альтернативой является параметрическая экспоненциальная линейная функция PeLU (Parametric Exponential Linear Unit), которым является обобщением ELU, а ее параметры формы функции обучаются в процессе тренировки сети.

PeLU расширяет ELU введением обучаемых параметров:

$$f'_{PeLU}(x) = \begin{cases} \dfrac{\beta}{\alpha}, & x > 0 \\ \beta\left(e^{\frac{x}{\alpha}} - 1\right), & x \le 0 \end{cases}, \tag{4}$$

где α – параметр масштабирования входа ($\alpha > 0$); β – параметр масштабирования выхода ($\beta > 0$).

Адаптивность параметров позволяет функции изменять форму в процессе обучения, повышая гибкость модели. Однако увеличение числа параметров усложняет оптимизацию. Как было отмечено ранее одним из решения вопроса с «умирающими нейронами» может быть не модификация ReLU, а ее замена на функцию активации которая будет лишена проблемы «умирающего нейрона». Такой Функцией является сигмоидальная функция Sigmoid. Даная функция является одной из классических активации, широко применявшаяся в ранних нейронных сетях [7].

Из-за того, что сигмоидальная функция имеет ненулевую производную при любых конечных значениях аргумента эта функция решает проблему с «умирающими нейронами».

Математическое выражение:

$$f_{\sigma}^{'}\left(x\right)=\frac{1}{1+e^{-x}}, \tag{9}$$

где e^{-x} – входное значение функции активации.

Таким образом, проведённое исследование функций активации и их производных, позволило выявить их ключевые отличия с точки зрения влияния на стабильность обучения, скорость сходимости и вероятность появления эффектов «затухающего градиента» и «умирающих нейронов».

Систематизация полученных теоретических результатов требует обобщения преимуществ и недостатков каждой функции активации, что даёт возможность сравнить их практическую эффективность и определить рациональность использования в глубоких нейронных сетях [9].

- ReLU обеспечивает высокую вычислительную эффективность, но подвержена деградации нейронов.
- ELU снижает вероятность потери активности.
- SeLU стабилизирует распределение активаций.
- PeLU повышает адаптивность модели.
- Sigmoid устойчива к умиранию нейронов, но вызывает затухание градиента.

В ходе исследования выполнен анализ функций активации ReLU, Sigmoid, ELU, SeLU и PeLU. Установлено, что наиболее рациональным выбором для задач компьютерного зрения в глубоких CNN является функция SeLU благодаря своему свойству самонормализации.

Список литературы:

1. Сацюк А.В. Мониторинг инфраструктуры на основе искусственного интеллекта // Автоматика, связь, информатика. – 2025. – № 9. – С. 32–34. – DOI: 10.62994/AT.2025.9.9.005. – EDN OXBPFP.
2. Сацюк А.В. Гибридные алгоритмы компьютерного зрения для прецизионного автономного управления БПЛА / А.В. Сацюк, Н.В. Володарец. – М. : Общество с ограниченной ответственностью «Научно-издательский центр ИНФРА-М», 2026. – 178 с. – ISBN 978-5-16-021492-4. – DOI: 10.12737/2227081. – EDN UTBINO.
3. Сацюк А.В. Особенности разработки беспилотных летательных аппаратов для отрасли железнодорожного транспорта / А.В. Сацюк, А.А. Воробьев // Сборник научных трудов Донецкого института железнодорожного транспорта. – 2023. – № 68. – С. 13–21. – EDN GSWIPU.
4. Сацюк А.В. Оптимизация архитектуры YOLOv8 для задач захвата объекта БПЛА: анализ компромисса между точностью, скоростью и вычислительными ресурсами // Вестник Ростовского государственного университета путей сообщения. – 2025. – № 2(98). – С. 35–42. – DOI: 10.46973/0201-727X_2025_2_35. – EDN LCAOOM.
5. Satsuk A.V. Acceleration of neural network for UAV operation in real time / A.V. Satsuk, N.V. Volodarets, D.I. Krupskiy // Vestnik Doneckoj Akademii Transporta. – 2025. – № 1. – P. 6–14. – EDN QVUIBK.
6. Сацюк А.В. Оценка эффективности алгоритмов YOLO для обнаружения объектов в реальном времени во встраиваемых системах беспилотных транспортных средств / А.В. Сацюк, Р.В. Белый, А.Е. Ищенко // Сборник научных трудов Донецкого института железнодорожного транспорта. – 2024. – № 4(75). – С. 73–82. – EDN OCNLZZ.
7. Снопкова И.А., Тауснев Д.А., Липинский Л.В. Исследование влияния функции активации на эффективность работы нейронной сети при различных подходах к обучению // Актуальные проблемы авиации и космонавтики. – 2020. – URL : https://cyberleninka.ru/article/n/issledovanie-vliyaniya-funktsii-aktivatsii-na-effektivnost-raboty-neyronnoy-seti-pri-razlichnyh-podhodah-k-obucheniyu (дата обращения: 19.02.2026).
8. Толмачев Сергей Владимирович, Гришина Любовь Сергеевна. Исследование влияния функции активации искусственной нейронной сети при решении задачи распознавания образов // Шаг в науку. – 2021. – № 4. – URL : https://cyberleninka.ru/article/n/issledovanie-vliyaniya-funktsii-aktivatsii-iskusstvennoy-neyronnoy-seti-pri-reshenii-zadachi-raspoznavaniya-obrazov (дата обращения: 19.02.2026).

9. Кашко Василий Васильевич, Олейникова Светлана Александровна. Анализ сходимости и устойчивости алгоритма глубокого обучения с подкреплением исполнитель-критик A2C // Вестник ВГТУ. – 2025. – № 1. – URL : https://cyberleninka.ru/article/n/analiz-shodimosti-i-ustoychivosti-algoritma-glubokogo-obucheniya-s-podkrepleniem-ispolnitel-kritik-a2c (дата обращения: 19.02.2026).

List of references:

1. Satsyuk A.V. Infrastructure Monitoring Based on Artificial Intelligence // Automation, Communications, Informatics. – 2025. – № 9. – P. 32–34. – DOI: 10.62994/AT.2025.9.9.005. – EDN OXBPFP.
2. Satsyuk A.V. Hybrid Computer Vision Algorithms for Precision Autonomous Control of UAVs / A.V. Satsyuk, N.V. Volodarets. – M. : Limited Liability Company «Scientific Publishing Center INFRA-M», 2026. – 178 p. – ISBN 978-5-16-021492-4. – DOI: 10.12737/2227081. – EDN UTBINO.
3. Satsuk A.V. Features of the development of unmanned aerial vehicles for the railway transport industry / A.V. Satsuk, A.A. Vorobyov // Collection of scientific papers of the Donetsk Institute of Railway Transport. – 2023. – № 68. – P. 13–21. – EDN GSWIPU.
4. Satsuk A.V. Optimization of the YOLOv8 architecture for UAV object capture tasks: analysis of the trade-off between accuracy, speed and computing resources // Bulletin of the Rostov State University of Railway Engineering. – 2025. – № 2 (98). – P. 35–42. – DOI: 10.46973/0201-727X_2025_2_35. – EDN LCAOOM.
5. Satsuk A.V. Acceleration of neural network for UAV operation in real time / A.V. Satsuk, N.V. Volodarets, D.I. Krupskiy // Vestnik Donetsk Academy of Transporta. – 2025. – № 1. – P. 6–14. – EDN QVUIBK.
6. Satsuk A.V. Evaluation of the effectiveness of YOLO algorithms for real-time object detection in embedded systems of unmanned vehicles / A.V. Satsuk, R.V. Belyy, A.E. Ishchenko // Collection of scientific papers of the Donetsk Institute of Railway Transport. – 2024. – № 4(75). – P. 73–82. – EDN OCNLZZ.
7. Snopkova I.A., Tausnev D.A., Lipinsky L.V. Study of the influence of the activation function on the efficiency of a neural network with different training approaches // Actual problems of aviation and cosmonautics. – 2020. – URL: https://cyberleninka.ru/article/n/issledovanie-vliyaniya-funktsii-aktivatsii-na-effektivnost-raboty-neyronnoy-seti-pri-razlichnyh-podhodah-k-obucheniyu (date of access: 19.02.2026).
8. Tolmachev Sergey Vladimirovich, Grishina Lyubov Sergeevna. Study of the influence of the activation function of an artificial neural network in solving a pattern recognition problem // Step into Science. – 2021. – № 4. – URL: https://cyberleninka.ru/article/n/issledovanie-vliyaniya-funktsii-aktivatsii-iskusstvennoy-neyronnoy-seti-pri-reshenii-zadachi-raspoznavaniya-obrazov (date of access: 19.02.2026).
9. Vasily Vasilyevich Kashko, Svetlana Aleksandrovna Oleynikova. Analysis of convergence and stability of the A2C performer-critic deep reinforcement learning algorithm // VSTU Bulletin. – 2025. – № 1. – URL: https://cyberleninka.ru/article/n/analiz-shodimosti-i-ustoychivosti-algoritma-glubokogo-obucheniya-s-podkrepleniem-ispolnitel-kritik-a2c (date of access: 19.02.2026).