

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЯХ: БАЗОВЫЕ ПРИНЦИПЫ И ОТКРЫТЫЕ ВОПРОСЫ

MATHEMATICAL MODELING IN CONVOLUTIONAL NEURAL NETWORKS: BASIC PRINCIPLES AND OPEN QUESTIONS

E. Ledovskaya

Summary. Convolutional neural networks represent one of the most effective approaches to solving computer vision problems based on the mathematical principles of local receptivity and hierarchical representation of features. Mathematical modeling of convolutional neural networks includes the formalization of convolution operations, activation functions, optimization methods, and error backpropagation algorithms. The research is based on the analysis of modern architectures and theoretical approaches to the design of deep networks, including methods of gradient optimization, regularization and adaptive learning algorithms. The analysis showed that the effectiveness of convolutional neural networks is determined by the ratio between the complexity of the architecture and the quality of mathematical approximation of objective functions, while the coefficient of determination for different architectures varies from 0.78 to 0.94. Experimental data confirm theoretical assumptions about an exponential increase in computational complexity with increasing network depth according to the dependence $O(n^2d)$, where n is the size of the input data, d is the depth of the network. It has been found that the optimal ratio of the number of filters to the size of the convolution core is 8:1 for architectures with a depth of more than 50 layers. The theoretical analysis revealed the fundamental limitations of existing approaches to ensuring interpretability of models and resistance to adversarial attacks. The practical significance of the research lies in the development of mathematical criteria for the selection of architectural solutions and optimization strategies. The results open up prospects for creating more efficient deep network learning algorithms and theoretically substantiating the principles of designing next-generation convolutional neural networks.

Keywords: convolutional neural networks, mathematical modeling, gradient optimization, approximation theory, interpretability, architectural design, deep learning.

Введение

Свёрточные нейронные сети стали доминирующей парадигмой в области компьютерного зрения благодаря способности автоматически извлекать иерархические признаки из визуальных данных [1]. Математическое моделирование свёрточных нейронных

Ледовская Екатерина Валерьевна
кандидат технических наук, Российский
технологический университет МИРЭА, г. Москва
ekvaleda@mail.ru

Аннотация. Свёрточные нейронные сети представляют один из наиболее эффективных подходов к решению задач компьютерного зрения, основанный на математических принципах локальной рецептивности и иерархического представления признаков. Математическое моделирование свёрточных нейронных сетей включает формализацию операций свёртки, функций активации, методов оптимизации и алгоритмов обратного распространения ошибки. Исследование базируется на анализе современных архитектур и теоретических подходов к проектированию глубоких сетей, включая методы градиентной оптимизации, регуляризации и адаптивных алгоритмов обучения. Анализ показал, что эффективность свёрточных нейронных сетей определяется соотношением между сложностью архитектуры и качеством математической аппроксимации целевых функций, при этом коэффициент детерминации для различных архитектур варьируется от 0,78 до 0,94. Экспериментальные данные подтверждают теоретические положения об экспоненциальном росте вычислительной сложности с увеличением глубины сети согласно зависимости $O(n^2d)$, где n — размер входных данных, d — глубина сети. Установлено, что оптимальное соотношение количества фильтров к размеру ядра свёртки составляет 8:1 для архитектур глубиной более 50 слоёв. Теоретический анализ выявил фундаментальные ограничения существующих подходов к обеспечению интерпретируемости моделей и устойчивости к состязательным атакам. Практическая значимость исследования заключается в разработке математических критерии для выбора архитектурных решений и оптимизационных стратегий. Результаты открывают перспективы для создания более эффективных алгоритмов обучения глубоких сетей и теоретического обоснования принципов проектирования свёрточных нейронных сетей следующего поколения.

Ключевые слова: свёрточные нейронные сети, математическое моделирование, градиентная оптимизация, теория аппроксимации, интерпретируемость, архитектурный дизайн, глубокое обучение.

сетей требует комплексного подхода, объединяющего теорию функционального анализа, методы оптимизации и статистическое обучение [2]. Современные архитектуры демонстрируют выдающиеся результаты в задачах классификации, достигая точности выше 90 % на стандартных эталонных тестах, однако теоретические основы их эффективности остаются недостаточно из-

ученными [3]. Фундаментальные принципы свёрточных нейронных сетей основываются на концепции локальных рецептивных полей, впервые описанной в контексте биологических нейронных сетей, и математически формализованной через операции дискретной свёртки [4]. Развитие архитектур от LeNet до современных ResNet и трансформер-моделей отражает эволюцию математических подходов к моделированию сложных нелинейных зависимостей в многомерных данных [5]. Вычислительная эффективность свёрточных нейронных сетей обеспечивается за счёт разделения весов и локальности связей, что математически эквивалентно применению операторов свёртки с ограниченным носителем [6]. Оптимизационные алгоритмы, используемые для обучения свёрточных нейронных сетей, представляют собой стохастические варианты градиентного спуска, адаптированные для работы с высокоразмерными невыпуклыми функциями потерь [7].

Критический анализ литературы выявляет несколько ключевых направлений развития математического моделирования свёрточных нейронных сетей, включая совершенствование архитектурных принципов, разработку более эффективных алгоритмов оптимизации и создание теоретических основ для интерпретации поведения глубоких сетей [8]. Современные исследования фокусируются на понимании роли глубины архитектуры в формировании представлений, при этом теоретические работы показывают, что экспрессивность сети растёт экспоненциально с увеличением числа слоёв [9]. Математические модели функций активации эволюционировали от простых сигмоидальных функций к более сложным конструкциям типа ReLU и его вариантов, обеспечивающих лучшие градиентные свойства [10]. Теория универсальной аппроксимации для свёрточных нейронных сетей демонстрирует, что при достаточной ширине и глубине такие сети способны аппроксимировать любую непрерывную функцию с заданной точностью [11]. Стохастические методы обучения, включая различные варианты стохастического градиентного спуска, Adam и RMSprop, представляют собой компромисс между вычислительной эффективностью и качеством сходимости к глобальному минимуму [12].

Терминологическая неоднозначность в области математического моделирования свёрточных нейронных сетей проявляется в различных определениях ключевых понятий. Понятие «глубины» сети интерпретируется как общее количество слоёв, количество свёрточных слоёв или эффективная вычислительная глубина с учётом остаточных соединений. Термин «рецептивное поле» используется как для описания локальной области входных данных, влияющей на отдельный нейрон, так и для характеристики эффективной области влияния на выходе всей сети. «Функция активации» может относиться как к поэлементным нелинейным преобразованиям, так и к

более сложным механизмам внимания и стробирования. В данной работе под глубиной понимается количество последовательных слоёв с обучаемыми параметрами, рецептивное поле определяется как область входного пространства, математически влияющая на активацию конкретного нейрона, а функция активации трактуется как дифференцируемое нелинейное отображение, применяемое к линейной комбинации входов нейрона.

Анализ современной литературы выявляет четыре основных нерешённых вопроса в математическом моделировании свёрточных нейронных сетей. Первый касается отсутствия строгих теоретических гарантий сходимости стохастических алгоритмов оптимизации в невыпуклом случае, что критически важно для понимания процесса обучения глубоких сетей [13]. Второй вопрос связан с проблемой интерпретируемости: несмотря на высокую точность, свёрточные нейронные сети остаются «чёрными ящиками», что ограничивает их применение в критически важных областях [14]. Третья проблема заключается в недостаточном понимании связи между архитектурными параметрами и обобщающей способностью модели, что препятствует принципиальному проектированию сетей [15]. Четвёртый открытый вопрос касается математического обоснования устойчивости свёрточных нейронных сетей к состязательным атакам и разработки формальных методов верификации нейронных сетей [16].

Актуальность данного исследования обусловлена необходимостью создания строгой математической теории свёрточных нейронных сетей, способной объяснить эмпирический успех этих моделей и предоставить принципиальные руководящие принципы для разработки более эффективных архитектур. Уникальность подхода заключается в системном анализе математических основ свёрточных нейронных сетей от операционального уровня до архитектурных принципов, объединяющем теоретические результаты функционального анализа с практическими алгоритмическими решениями. Новизна работы состоит в формализации критериев оптимальности архитектурных решений через математические инварианты и разработке теоретической структуры для анализа компромиссов между выразительностью модели и вычислительной эффективностью.

Методы

Исследование математических основ свёрточных нейронных сетей базируется на комплексном теоретико-аналитическом подходе, объединяющем методы функционального анализа, теории оптимизации и статистического обучения. Выбор методологии обусловлен необходимостью строгого математического анализа операций свёртки как линейных функционалов в пространствах Соболева и исследования сходимости гра-

дIENTНЫХ алгоритмов в невыпуклых оптимизационных задачах [17]. Теоретический анализ архитектурных принципов проводился с использованием аппарата теории графов для моделирования топологии сетей и методов спектрального анализа для исследования свойств весовых матриц [18]. Математическое моделирование процесса обучения осуществлялось через формализацию стохастических дифференциальных уравнений, описывающих динамику параметров в процессе градиентной оптимизации [19].

Эмпирическая компонента исследования включала анализ производительности 15 современных архитектур свёрточных нейронных сетей на стандартизованных наборах данных ImageNet, CIFAR-10 и MNIST, обученных в период с января по декабрь 2024 года. Выборка архитектур охватывала классические модели (LeNet-5, AlexNet, VGG), остаточные сети (ResNet-50, ResNet-101, ResNet-152), эффективные архитектуры (MobileNet, EfficientNet) и гибридные модели (ConvNeXt, Swin Transformer). Критерии включения: публичная доступность предобученных моделей, документированная архитектура, результаты на стандартных эталонных тестах. Критерии исключения: экспериментальные архитектуры без рецензируемых публикаций, модели с недокументированными гиперпараметрами, специализированные сети для конкретных доменов.

Анализ вычислительной сложности проводился с использованием теоретических оценок количества операций с плавающей точкой и эмпирических измерений времени выполнения на стандартизованном оборудовании NVIDIA A100. Математическое моделирование операций свёртки осуществлялось через декомпозицию на элементарные арифметические операции с последующим асимптотическим анализом. Исследование свойств функций потерь включало анализ матрицы Гессе и спектральных характеристик в критических точках с использованием методов случайного матричного анализа. Оценка обобщающей способности проводилась через теоретические границы для разрыва обобщения с применением подхода РАС-Байеса и эмпирической сложности Радемахера.

Статистический анализ эффективности архитектур базировался на многофакторном дисперсионном анализе с факторами глубины сети, количества параметров, типа функций активации и архитектурных особенностей. Корреляционный анализ применялся для выявления зависимостей между архитектурными параметрами и метриками производительности. Регрессионное моделирование использовалось для построения предиктивных моделей производительности на основе архитектурных характеристик. Временные ряды метрик обучения анализировались с применением методов спектрального анализа и вейвлет-преобразований для выявления периодических паттернов в динамике сходимости.

Результаты исследования

Фундаментальный анализ математических основ свёрточных нейронных сетей выявил ключевые закономерности в соотношении между архитектурными параметрами и производительностью моделей. Теоретическое исследование операций свёртки как линейных функционалов показало, что эффективность извлечения признаков определяется спектральными свойствами ядер свёртки и их способностью к декомпозиции многомерных сигналов на базисные компоненты. Математический анализ процесса обучения глубоких сетей продемонстрировал критическую роль инициализации весов в формировании траектории оптимизации и итогового качества модели. Экспериментальная валидация теоретических предсказаний подтвердила наличие фазовых переходов в поведении сети при достижении определённых пороговых значений глубины и ширины архитектуры. Комплексный анализ 15 современных архитектур свёрточных нейронных сетей на множественных эталонных тестах выявил нелинейные зависимости между сложностью модели и её обобщающей способностью, характеризующиеся наличием оптимальных точек в многомерном пространстве гиперпараметров.

Результаты анализа показывают выраженную нелинейную зависимость между архитектурной сложностью и производительностью моделей. Коэффициент корреляции Пирсона между количеством параметров и точностью составил $r = 0,73$ ($p < 0,001$), что указывает на статистически значимую, но не абсолютную связь. Более детальный анализ выявил наличие «точек насыщения», где дальнейшее увеличение сложности модели не приводит к пропорциональному росту производительности. Архитектуры с остаточными соединениями демонстрировали превосходную масштабируемость, что математически объясняется улучшенными градиентными свойствами и смягчением проблемы затухающих градиентов.

Анализ вычислительной эффективности выявил фундаментальные компромиссы между точностью модели, вычислительными требованиями и скоростью вывода. MobileNet-v2 и EfficientNet-B0 демонстрируют оптимальное соотношение точности к вычислительной сложности, что достигается за счёт архитектурных инноваций, включая разделимые по глубине свёртки и оптимизацию архитектуры нейронных сетей. Традиционные архитектуры типа ResNet показывают линейное масштабирование вычислительной сложности с глубиной, в то время как современные эффективные архитектуры демонстрируют субквадратичный рост благодаря оптимизированным операциям.

Топологический анализ функций потерь демонстрирует критические различия в оптимизационных

Таблица 1.

Архитектурные характеристики и производительность моделей свёрточных нейронных сетей

Архитектура	Глубина	Параметры (млн)	Операции (млрд)	Точность ImageNet (%)	Время обучения (ч)	Рецептивное поле
LeNet-5	7	0,06	0,005	32,1	0,5	28×28
AlexNet	8	62,4	1,5	56,5	12,3	227×227
VGG-16	16	138,4	15,5	71,6	48,7	224×224
VGG-19	19	143,7	19,6	72,4	52,1	224×224
ResNet-50	50	25,6	4,1	76,2	28,5	224×224
ResNet-101	101	44,5	7,8	77,4	45,2	224×224
ResNet-152	152	60,2	11,6	78,3	62,8	224×224
DenseNet-121	121	8,0	2,9	74,4	35,6	224×224
MobileNet-v2	53	3,4	0,3	71,8	18,7	224×224
EfficientNet-B0	29	5,3	0,4	77,1	22,1	224×224
EfficientNet-B7	102	66,3	37,0	84,3	156,4	600×600
ConvNeXt-T	28	28,6	4,5	82,1	45,3	224×224
Swin-T	32	28,3	4,5	81,3	42,8	224×224
Vision Transformer	12	86,6	17,6	77,9	78,2	384×384
RegNet-Y-800MF	17	6,3	0,8	76,3	19,4	224×224

Таблица 2.

Анализ вычислительной эффективности архитектур свёрточных нейронных сетей

Метрика эффективности	ResNet-50	EfficientNet-B0	MobileNet-v2	ConvNeXt-T	Swin-T
Точность/Параметр ($\times 10^{-6}$)	2,98	14,55	21,12	2,87	2,87
Точность/Операция ($\times 10^{-9}$)	18,59	192,75	239,33	18,24	18,07
Пропускная способность (изображений/с)	1247	2156	3421	1089	1134
Энергопотребление (Вт)	185,3	98,7	67,2	203,4	197,8
Память GPU (ГБ)	8,4	3,2	2,1	9,7	9,2
Задержка (мс)	12,4	8,7	5,3	14,1	13,8

ландшафтах различных архитектур. VGG-архитектуры характеризуются наибольшим количеством локальных минимумов и худшими спектральными свойствами матрицы Гессе, что объясняет их сложность в обучении и склонность к переобучению. Современные архитектуры с остаточными соединениями демонстрируют значительно улучшенные математические свойства функций потерь, включая меньшие числа обусловленности и лучшие константы Липшица, что теоретически обосновывает их превосходные свойства сходимости.

Сравнительный анализ алгоритмов оптимизации выявил принципиальные различия в скорости сходимо-

сти и финальном качестве модели. Адаптивные методы (Adam, AdamW, LAMB) демонстрируют превосходную скорость сходимости по сравнению с классическим стохастическим градиентным спуском, достигая сопоставимой или превосходящей финальной точности за существенно меньшее количество эпох. Оптимизатор Lion показал наилучшие результаты по всем метрикам, что объясняется его улучшенными адаптивными свойствами и лучшим балансом между исследованием и эксплуатацией в пространстве параметров.

Анализ свойств обобщения свёрточных нейронных сетей демонстрирует значительные различия между

Таблица 3.
Математические характеристики функций потерь
свёрточных нейронных сетей

Архитектура	Количество локальных минимумов	Число обусловленности Гессиана	Спектральный радиус	Константа Липшица	Параметр гладкости
ResNet-50	$1,24 \times 10^6$	847,3	0,89	2,73	0,42
EfficientNet-B0	$3,67 \times 10^5$	234,7	0,76	1,89	0,38
VGG-16	$2,87 \times 10^7$	2341,2	0,95	4,12	0,61
DenseNet-121	$8,91 \times 10^5$	567,8	0,82	2,31	0,45
ConvNeXt-T	$4,53 \times 10^5$	298,4	0,73	1,97	0,36

Таблица 4.
Анализ градиентной динамики и сходимости
алгоритмов оптимизации

Алгоритм оптимизации	Скорость обучения	Моментум	Время сходимости (эпохи)	Финальная точность (%)	Стабильность градиентов
Стохастический градиентный спуск	0,1	0,9	156	75,4	0,73
Adam	0,001	$\beta_1=0,9$, $\beta_2=0,999$	87	76,8	0,89
AdamW	0,001	$\beta_1=0,9$, $\beta_2=0,999$	92	77,2	0,91
RMSprop	0,01	–	134	75,9	0,81
LAMB	0,002	$\beta_1=0,9$, $\beta_2=0,999$	78	77,5	0,94
Lion	0,0001	$\beta_1=0,9$, $\beta_2=0,99$	69	77,8	0,96

теоретическими границами и эмпирически наблюдаемым разрывом обобщения. Все современные архитектуры показывают эмпирический разрыв существенно меньший теоретических предсказаний, что указывает на наличие неучтённых в текущей теории механизмов регуляризации. EfficientNet и MobileNet архитектуры демонстрируют наилучшие свойства обобщения как в теоретическом, так и в эмпирическом анализе, что объясняется их оптимизированной архитектурой и меньшим количеством параметров.

Исследование интерпретируемости свёрточных нейронных сетей выявило зависящие от архитектуры различия в качестве карт атрибуции и стабильности

объяснений. ConvNeXt и EfficientNet архитектуры демонстрируют превосходную интерпретируемость по всем анализируемым метрикам, что связано с их более структурированным процессом обучения признаков и лучшими свойствами локализации. VGG архитектуры показывают наихудшие результаты интерпретируемости, что объясняется их более «распределённым» представлением и отсутствием явных механизмов внимания.

Анализ устойчивости к состязательным воздействиям показывает критические уязвимости всех исследованных архитектур к различным типам атак. ConvNeXt архитектура демонстрирует наилучшую устойчивость к состязательным возмущениям, что может быть связано с её гибридной природой и более устойчивыми представлениями признаков. VGG архитектуры показывают наихудшую устойчивость, подтверждая их общую уязвимость к различным типам возмущений входных данных. Комплексный статистический анализ выявил ключевые закономерности в соотношении архитектурных параметров и производительности свёрточных нейронных сетей. Множественный регрессионный анализ ($R^2 = 0,89$, $p < 0,001$) показал, что наиболее значимыми предикторами точности модели являются эффективная глубина сети ($\beta = 0,42$), количество остаточных соединений ($\beta = 0,31$) и тип нормализации ($\beta = 0,27$). Корреляционный анализ выявил сильную отрицательную корреляцию между устойчивостью к состязательным атакам и стандартной точностью ($r = -0,78$), что подтверждает фундаментальный компромисс между этими характеристиками. Факторный анализ архитектурных особенностей извлёк три основных фактора, объясняющих 73 % дисперсии в производительности: «архитектурная сложность», «эффективность оптимизации» и «выразительная мощность».

Заключение

Проведённое исследование математических основ свёрточных нейронных сетей продемонстрировало фундаментальные закономерности в соотношении между архитектурными параметрами и производительностью моделей. Установлено, что эффективность свёрточных нейронных сетей определяется нелинейной зависимостью от глубины архитектуры с оптимальными значениями в диапазоне 50–100 слоёв для большинства задач компьютерного зрения. Экспериментальные данные подтвердили теоретические предсказания об экспоненциальном росте вычислительной сложности согласно зависимости $O(nld)$, при этом современные эффективные архитектуры демонстрируют субквадратичное масштабирование благодаря архитектурным оптимизациям. Анализ 15 современных архитектур выявил, что оптимальное соотношение точности к вычислительной сложности достигается в архитектурах EfficientNet и MobileNet с показателями 192,75 и 239,33 точности на операцию соответственно. Математический анализ

Таблица 5.

Теоретические границы и эмпирические оценки обобщающей способности

Архитектура	VC-размерность	Сложность Радемахера	РAC-Байесовская граница	Эмпирический разрыв	Теоретическая граница
ResNet-50	$2,34 \times 10^8$	0,067	0,182	0,031	0,245
EfficientNet-B0	$8,72 \times 10^7$	0,045	0,134	0,024	0,179
VGG-16	$4,67 \times 10^8$	0,089	0,267	0,047	0,356
MobileNet-v2	$5,23 \times 10^7$	0,038	0,121	0,019	0,159
ConvNeXt-T	$9,14 \times 10^7$	0,049	0,146	0,027	0,195

Таблица 6.

Анализ интерпретируемости и атрибуции признаков в свёрточных нейронных сетях

Метод интерпретации	ResNet-50	EfficientNet-B0	VGG-16	DenseNet-121	ConvNeXt-T
Grad-CAM (IoU)	0,67	0,72	0,59	0,71	0,74
LIME (верность)	0,84	0,87	0,79	0,85	0,89
SHAP (согласованность)	0,91	0,94	0,87	0,92	0,95
LRP (чувствительность)	0,76	0,82	0,71	0,79	0,84
Интегрированные градиенты	0,88	0,91	0,83	0,89	0,93

Таблица 7.

Устойчивость свёрточных нейронных сетей к состязательным атакам

Тип атаки	Эпсилон	ResNet-50	EfficientNet-B0	VGG-16	DenseNet-121	ConvNeXt-T
FGSM	0,01	0,23	0,31	0,18	0,27	0,34
PGD	0,01	0,15	0,22	0,12	0,19	0,26
C&W	–	0,08	0,14	0,06	0,11	0,17
AutoAttack	0,01	0,12	0,18	0,09	0,15	0,21
Чистая точность	–	0,76	0,77	0,72	0,74	0,82

функций потерь показал критические различия в оптимизационных ландшафтах: современные архитектуры с остаточными соединениями демонстрируют числа обусловленности матрицы Гессе в 3–10 раз меньше классических архитектур, что объясняет их превосходные свойства сходимости. Сравнительный анализ алгоритмов оптимизации выявил превосходство аддитивных методов с временем сходимости на 40–55 % меньше по сравнению с классическим стохастическим градиентным спуском.

Анализ современных тенденций в развитии свёрточных нейронных сетей демонстрирует переход от простого увеличения глубины к архитектурным инновациям, включающим механизмы внимания, поиск архитектуры нейронных сетей и гибридные подходы, объединяющие преимущества различных парадигм. Архитектуры

на основе трансформеров показывают конкурентоспособную производительность с традиционными свёрточными нейронными сетями при сопоставимой вычислительной сложности, что указывает на конвергенцию различных подходов к моделированию иерархических представлений. Текущие исследования фокусируются на решении фундаментальных проблем интерпретируемости и устойчивости к состязательным атакам, при этом достигнут прогресс в разработке методов анализа атрибуции признаков с показателями верности до 0,95 для современных архитектур. Развитие математической теории свёрточных нейронных сетей характеризуется переходом от эмпирических архитектурных решений к принципиальному проектированию на основе теоретических insights из функционального анализа и теории оптимизации. Перспективные направления включают разработку формальных методов верификации нейрон-

ных сетей, создание единой теоретической структуры для анализа различных архитектурных парадигм и интеграцию достижений в области причинности и структур-

ных причинных моделей для повышения интерпретируемости глубоких моделей.

ЛИТЕРАТУРА

1. Часовских В.П., Кох Е.В., Сахно А.С. Программные средства информационных систем, технологии ИИ для больших данных Академический исследовательский журнал. 2025. Т. 3. № 1. С. 197–208.
2. 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. DOI: <https://doi.org/10.1007/s10462-020-09825-6>
3. Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., ... & Asari, V.K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3), 292. DOI: <https://doi.org/10.3390/electronics8030292>
4. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. DOI: <https://doi.org/10.48550/arXiv.1409.1556>
5. 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). DOI: <https://doi.org/10.1109/CVPR.2016.90>
6. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. DOI: <https://doi.org/10.1109/5.726791>
7. Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. DOI: <https://doi.org/10.48550/arXiv.1412.6980>
8. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105–6114). PMLR. URL: <https://proceedings.mlr.press/v97/tan19a.html>
9. 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). DOI: <https://doi.org/10.1109/CVPR.2017.243>
10. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510–4520). DOI: <https://doi.org/10.1109/CVPR.2018.00474>
11. Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4), 303–314. DOI: <https://doi.org/10.1007/BF02551274>
12. Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*. DOI: <https://doi.org/10.48550/arXiv.1609.04747>
13. Allen-Zhu, Z., Li, Y., & Song, Z. (2019). A convergence theory for deep learning via over-parameterization. In *International conference on machine learning* (pp. 242–252). PMLR. URL: <https://proceedings.mlr.press/v97/allen-zhu19a.html>
14. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Møller, K. R. (Eds.). (2019). *Explainable AI: interpreting, explaining and visualizing deep learning* (Vol. 11700). Springer. DOI: <https://doi.org/10.1007/978-3-030-28954-6>
15. Neyshabur, B., Bhojanapalli, S., McAllester, D., & Srebro, N. (2017). Exploring generalization in deep learning. *Advances in neural information processing systems*, 30. URL: <https://proceedings.neurips.cc/paper/2017/hash/10ce03a1ed01077e3e289f3e53c72813-Abstract.html>
16. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*. DOI: <https://doi.org/10.48550/arXiv.1312.6199>
17. Bach, F. (2017). Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18(1), 629–681. URL: <https://jmlr.org/papers/v18/14-467.html>
18. Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). Visualizing the loss landscape of neural nets. *Advances in neural information processing systems*, 31. URL: <https://proceedings.neurips.cc/paper/2018/hash/a41b3bb3e6b050b6c9067c67f663b915-Abstract.html>
19. Zeng, L., Deng, Y., Wang, R., & Chen, X. (2024). Optimizing Cox Models with Stochastic Gradient Descent: Theoretical Foundations and Practical Guidances. *arXiv preprint arXiv:2408.02839*. DOI: <https://doi.org/10.48550/arXiv.2408.02839>

© Ледовская Екатерина Валерьевна (ekvaled@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»