

Федеральное государственное автономное образовательное учреждение  
высшего образования «Национальный исследовательский университет  
«Высшая школа экономики»

На правах рукописи

ПАВЛОВ СТАНИСЛАВ ВЛАДИМИРОВИЧ

**ОБОБЩЕНИЕ НЕЙРОННЫХ СЕТЕЙ НА АЛГЕБРУ ДУАЛЬНЫХ ЧИСЕЛ**

РЕЗЮМЕ

диссертации на соискание ученой степени

кандидата компьютерных наук

Научный руководитель:

доктор физико-математических наук,

профессор Калягин Валерий Александрович

Нижний Новгород — 2024

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Нейронные сети в настоящее время применяются в задачах классификации, детекции, семантической сегментации и обработки естественных языков. Однако, проблемы ограниченности точности и баланса между скоростью исполнения и точностью алгоритмов существенно ограничивают сферу эффективности их применения. Современные подходы обобщения нейронных сетей на гиперкомплексные алгебры направлены на решение этих проблем [1], [2].

**Актуальность темы** обусловлена возрастающей популярностью направления по обобщению нейронных сетей на гиперкомплексные алгебры. С одной стороны, это вызвано тем, что исходные данные часто представляются в комплексной форме, с другой стороны, целый ряд новых исследований демонстрируют превосходство использования комплексных нейронных сетей, по сравнению с реальными, в том числе, большую точность [3] и лучшую сходимость [4].

Комплексные нейронные сети находят всё большее применение в различных прикладных и исследовательских задачах [5]. Решаются такие задачи, как обработка радиосигналов, обработка изображений и компьютерное зрение, обработка и анализ аудио сигналов, обработка сигналов с радаров и санаров, криптография, предсказание временных рядов, ассоциативная память, предсказание ветра, робототехника, контроль транспортных потоков, обнаружение спама, предсказания в агрокультуре и другие.

Автор В. Widrow одним из первых предложил комплексный LMS алгоритм и показал его эффективность в сравнении с вещественными [6]. В работе А. Hirose по обработке радиосигналов [7] сравниваются обобщающие характеристики

комплексных и вещественных нейронных сетей с точки зрения когерентности обрабатываемых сигналов. Исследуется задача аппроксимации функций (интерполяцию временных сигналов). Моделирование и реальные эксперименты показывают, что комплексная нейронная сеть с функцией активации амплитудно-фазового типа показывают меньшую погрешность обобщения, чем реально значимые сети, такие как бивариатные вещественные нейронные сети. Также следующие авторы внесли существенный вклад в применение идей комплексных нейронных сетей в обработке радио сигналов: Т. Kim, S. Scardapane, Y. Quan, M. Catelani, A. Marseet, I. Cha, S. Chen, D. Jianping, W. Gong, A. Uncini, M. Scarpiniti, R. Huang, M. Solazzi, N. Benvenuto, A. B. Suksmono, A. Hirose, Ю. Чистяков, А. Минин, J. Zhang, S. Liu, M. Peker, S. Hu, Y. Suzuki, T. Ding [8]-[34].

Существенный вклад в развитии комплексных сетей для задач компьютерного зрения внесли авторы: М. Arjovsky, С.-А. Popa, М. Matlacz, И. Н. Айзенберг, Р. Virtue, Е. Айзенберг, R. S. Zemel, С. Trabelsi, S. Amilia, М. Miyauchi, А. Hirose, Y. Liu, R. F. Olanrewaju, R. Hata, Y. Kominami, С.-А. Popa, L. Li, S. Gu [1], [35]-[57]. Комплексные глубокие нейронные сети некоторое время были ограничены в применении для задач компьютерного зрения из-за отсутствия необходимых строительных блоков. Знаковая работа автора С. Trabelsi [45] предоставляет ключевые компоненты для комплексных глубоких нейронных сетей и показывает их применение к свёрточным нейронным сетям и LSTM. С. Trabelsi предложил комплексные свёртки и несколько вариантов алгоритмов для комплексной пакетной нормализации, стратегии инициализации весов для комплексных нейронных сетей, а также показал преимущества по отношению к вещественным

аналогам в задачах компьютерного зрения. В работе автора S. Gu [57] предлагается комплексный аналог VGG, комплексный аналог полносвязного слоя и показывают преимущество такой архитектуры для задач распознавания, добиваясь лучшего качества в аналогичном классе архитектур для того времени.

Многие исследователи расширяли применение комплексных нейронных сетей на другие задачи: обработка и анализ аудио сигналов - C. Trabelsi, D. Hayakawa, M. Kataoka, M. Kinouchi, A. Y. H. Al-Nuaimi, Y.-S. Lee, C. S. Tay [45], [58-63], обработка сигналов с радаров и санаров - J. Gao, M. Wilmanski, И. Н. Айзенберг, К. Оуама, X. Yao [64]-[69], криптография - T. Dong [70], предсказание временных рядов – И. Н. Айзенберг [66], ассоциативная память - S. Jankowski, T. Miyajima [71-72], предсказание ветра - H. H. Cevik, T. Kitajima, D. P. Mandic [73-75], робототехника - Y. Maeda [76], контроль транспортных потоков - I. Nishikawa [77-78], обнаружение спама - J. Hu [79], предсказания в агрокультуре – И. Н. Айзенберг [37].

Эти потрясающие результаты вдохновили на дальнейшее обобщение нейронных сетей другие гиперкомплексные числа, в частности, на дуальные числа. Дуальные числа уже нашли применение в винтовой теории (F. M. Dimentberg) [80], также дуальные числа позволяют автоматически считать производную (A. Güneş Baydin, R. Kiran) [81-82]. Вместе с этим, известна только одна базовая попытка применения дуальный числах в нейронных сетях (Y. Okawa) [83], где использовали свойства дуальных чисел для входных данных. Данное направление заслуживает дальнейшего исследования.

**Целью диссертационной работы** является обобщение нейронных сетей на

алгебру дуальных чисел для достижения лучшего соотношения качество-скорость вычислений.

Для достижения поставленной цели были решены следующие **задачи**:

- 1) Разработать математическую основу нейронных сетей на гиперкомплексных алгебрах.
- 2) Разработать методологию построения нейронных сетей на гиперкомплексных алгебрах.
- 3) Проведение вычислительных экспериментов для демонстрации преимуществ нового подхода.

**Объектом исследования** являются нейронные сети.

**Предметом исследований** является математическое и алгоритмическое обеспечение обобщения нейронных сетей на гиперкомплексную алгебру.

**Научная новизна работы** состоит в следующем:

- 1) Определена формула гиперкомплексной нормы и алгоритм пакетной нормализации, основанный на этой норме. Также определена формула производной для функций гиперкомплексных переменных второго порядка.
- 2) Определены 15 операторов нейронных сетей на гиперкомплексной алгебре, включая основные (свёртка, линейный, групповая нормировка, пулинг, блок линейной ректификации), определены дуальные голоморфные операторы. Проведены вычислительные эксперименты по построению гиперкомплексных нейронных сетей. Разработана и процедура переноса знаний от вещественных сетей к гиперкомплексным.
- 3) Показаны преимущества (точность и скорость) разработанного подхода к

решению ряда задач (компьютерное зрение, обнаружение гравитационных волн и транскрипция музыки), за счёт выделения признаков дуального типа, которые не были рассмотрены ранее.

**Практическая ценность.** Показана возможность применения разработанного подхода к решению практических задач (приведены базовые реализации гиперкомплексных сетей, классические задачи компьютерного зрения, обнаружение гравитационных волн, задача по транскрипции музыки, а также улучшения с применением дуальных голоморфных нейронных сетей). Выполнена программная реализация в продукте с открытым доступом MindSpore.

**Внедрения результатов работы.** Результаты исследования внедрены в учебный процесс на кафедре прикладной математики и информатики. Гиперкомплексные операторы и сети внедрены в продукт с открытым доступом MindSpore.

**Методы исследования.** В работе используются современные методы машинного обучения, теории глубоких нейронных сетей, алгебры гиперкомплексных чисел.

**Основные положения, выносимые на защиту:**

- 1) Формула гиперкомплексной нормы и алгоритм пакетной нормализации, основанный на этой норме. Формула производной для функций в гиперкомплексных алгебрах второго порядка.
- 2) Разработка архитектуры нейронных сетей для гиперкомплексных алгебр. Разработка и программная реализация операторов нейронных сетей на гиперкомплексной алгебре.

- 3) Результаты по оценке эффективности различных стратегий переноса знаний от вещественных нейронных сетей к гиперкомплексным.
- 4) Эффективность (по скорости и точности) применения гиперкомплексных нейронных сетей в задачах классического компьютерного зрения, обнаружения гравитационных волн и транскрипции музыки и других.

**Достоверность результатов.** Достоверность результатов обеспечена корректной разработкой математического аппарата и проведением экспериментальных исследований.

**Апробация работы.** Основные положения и результаты диссертации доложены и обсуждены на следующих научно-технических конференциях и семинарах:

- 1) Международная конференция по продвинутому наблюдению на основе видео и сигналов, IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS, Мадрид, Испания, 29 ноября – 2 декабря 2022), доклад «Dual-valued Neural Networks».
- 2) Международная конференция по акустике, речи и обработке сигналов, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP, Остров Родос, Греция, с 4 по 10 июня 2023), доклад «Learning Properties of Holomorphic Neural Networks of Dual Variables».
- 3) Научный семинар в Applied AI Center, Skolkovo Institute of Science and Technology (12 ноября 2023), доклад «Обобщение нейронных сетей на алгебру дуальных чисел».
- 4) Расширенное заседание научного семинара лаборатории алгоритмов и технологий анализа сетевых структур НИУ ВШЭ (20 декабря 2023), доклад на

тему «Generalization of neural networks on the dual numbers algebra».

**Публикации.** По теме диссертации опубликовано 3 работы.

**Личный вклад:** автором разработана ключевая идея обобщения нейронных сетей на алгебру дуальных чисел, математический аппарат и алгоритмическая часть, проведены эксперименты и сделаны выводы.

Публикации повышенного уровня:

- 1) Pavlov, S.; Kozlov, D.; Bakulin, M.; Zuev, A.; Latyshev, A.; Beliaev, A. Generalization of Neural Networks on Second-Order Hypercomplex Numbers. Mathematics 2023, 11, 3973. <https://doi.org/10.3390/math11183973>. Scopus Q1 journal, Q2 по базе scimagojr.com. Личный вклад: автором разработана ключевая идея обобщения нейронных сетей на алгебру дуальных чисел, разработаны концепции и подходы, математический аппарат (включающий дуальную производную и градиент) и алгоритмическая часть (включающая дуальные архитектуры и операторы нейронной сетей), проведены эксперименты по обучению, сделаны оптимизации и выводы.

Публикации стандартного уровня:

- 2) Dmitry Kozlov; Stanislav Pavlov; Alexander Zuev; Mikhail Bakulin; Mariya Krylova, Igor Kharchikov. Dual-valued Neural Networks. 2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS, Madrid, Spain). DOI: 10.1109/AVSS56176.2022.9959227. Core B, IEEE Xplore, WoS. Личный вклад: автором разработана ключевая идея обобщения нейронных сетей на алгебру дуальных чисел, математический аппарат (включающий дуальную норму) и алгоритмическая часть (включающая

дуальные операторы нейронной сетей), проведены эксперименты по обучению и сделаны выводы.

- 3) Dmitry Kozlov; Mikhail Bakulin; Stanislav Pavlov; Aleksandr Zuev; Mariya Krylova, Igor Kharchikov. Learning Properties of Holomorphic Neural Networks of Dual Variables. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP, Rhodes Island, Greece). DOI: 10.1109/ICASSP49357.2023.10095457. Core B, IEEE Xplore, WoS, 13я конференция по импакт-фактору в мире <https://research.com/conference-rankings/computer-science>. Личный вклад: автором разработана ключевая идея обобщения нейронных сетей на алгебру дуальных чисел, понятие голоморфных дуальных сетей, математический аппарат (включающий понятие дуальных голоморфных функций) и алгоритмическая часть (включающая дуальные голоморфные операторы нейронной сетей), проведены эксперименты по обучению, сделаны оптимизации и выводы.

**Структура и объем работы.** Диссертация состоит из введения, двух глав, заключения, списка литературы и приложений. Общий объем работы 97 страниц текста, содержащего 14 рисунков и схем, 10 таблиц и 58 формул. Список литературы содержит 118 источников.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** обоснована актуальность диссертационной работы, сформулированы цель и задачи исследования, объяснена научная новизна и практическая ценность полученных результатов, представлены выносимые на защиту положения, дана общая характеристика работы.

**В первой главе** описана теория дуальных чисел в нейронных сетях: поставлена задача, описано обобщение нейронных сетей ко всем гиперкомплексным числам второго порядка, представление данных в гиперкомплексной алгебре, гиперкомплексные операции, норма гиперкомплексных чисел, гиперкомплексная пакетная нормализация, обратное распространение градиента функции потери, гиперкомплексный преобразователь.

В ходе исследования дуальных операторов и нейронных сетей автор обнаружил, что многие аспекты этой задачи очень близки (в некоторых случаях совпадают) для комплексных и дуальных чисел. Все эти алгебры (дуальная  $\mathbb{D}$ , комплексная  $\mathbb{C}$  и двойная или комплексно-расщеплённые  $\mathbb{S}$ ) являются гиперкомплексными второго порядка.

По определению множество дуальных чисел является коммутативным кольцом  $\langle \mathbb{D}, +, \times \rangle$ , которое является надмножеством поля действительных чисел  $\langle \mathbb{R}, +, \times \rangle$ , и имеет специальный элемент  $\varepsilon$  чтобы каждый элемент  $\mathbb{D}$  можно представить  $a + \varepsilon b$  для  $a, b \in \mathbb{R}$ , и  $\varepsilon^2 = 0$ ,  $\varepsilon \neq 0$ . Точно так же множество двойных чисел  $\langle \mathbb{S}, +, \times \rangle$  это суперсет  $\langle \mathbb{R}, +, \times \rangle$  со специальным элементом  $j$  так, что каждый элемент из  $\mathbb{S}$  может быть представлена как  $a + jb$  для  $a, b \in \mathbb{R}$ , и  $j^2 = 1$  (но  $j \notin \mathbb{R}$ ). Дуальные-, двойные-, and комплексные кольца являются изоморфными специальной алгебре Клиффорда включающей единицу,  $e_0$ ,  $e_1$ , и  $e_0e_1$ , где  $e_0^2 = -1$ ,  $e_1^2 = \sigma \in \{0, 1, -1\}$  и  $-e_0e_1 = e_1e_0 = \tau$ , что эквивалентно  $\varepsilon$ ,  $j$  или  $i$ , соответственно. В каждом случае,  $\tau^2 = \sigma$  по определению кольца соответствующих чисел. Это математическое единство позволяет нам обобщать все гиперкомплексные алгебры второго порядка и оперировать с ними одинаково.

Основные математические операции для этих чисел:

$$(x_1 + \tau y_1) \pm (x_2 + \tau y_2) = (x_1 \pm x_2) + \tau(y_1 \pm y_2),$$

$$(x_1 + \tau y_1) \cdot (x_2 + \tau y_2) = (x_1 x_2 + \sigma y_1 y_2) + \tau(x_1 y_2 + y_1 x_2),$$

$$\frac{x_1 + \tau y_1}{x_2 + \tau y_2} = \frac{x_1 x_2 - \sigma y_1 y_2}{x_2^2 - \sigma y_2^2} + \tau \frac{x_2 y_1 - x_1 y_2}{x_2^2 - \sigma y_2^2},$$

$$(x + \tau y)^* = x - \tau y.$$

Автор связывает мнимую часть исходного входа с дуальным компонентом

$$x + iy \Rightarrow x + \varepsilon y \text{ or } x + jy .$$

Чтобы объяснить свёртку в гиперкомплексной алгебре, автор использует матричное представление гиперкомплексных чисел второго порядка, которое использует вещественные числа. Известно, что алгебры комплексных, дуальных и двойных чисел  $u=x+\tau y$  изоморфны соответствующим алгебрам вещественных матриц второго порядка вида  $\begin{pmatrix} x & y \\ \sigma y & x \end{pmatrix}$ . Таким образом, свертка гиперкомплексного фильтра  $W = W_x + \tau W_y$  и гиперкомплексный вход  $u=x+\tau y$  можно выразить следующим образом:

$$\begin{aligned} W * u &= \begin{pmatrix} W_x & W_y \\ \sigma W_y & W_x \end{pmatrix} * \begin{pmatrix} x & y \\ \sigma y & x \end{pmatrix} \\ &= \begin{pmatrix} W_x * x + \sigma W_y * y & W_x * y + W_y * x \\ \sigma(W_x * y + W_y * x) & W_x * x + \sigma W_y * y \end{pmatrix}. \end{aligned}$$

Линейный слой. Чтобы обобщить линейные слои для двумерных алгебр, автор использует матричное представление гиперкомплексных чисел второго порядка. Линейный слой с гиперкомплексными входами и весами эквивалентен суперпозиции вещественных линейных слоев:

$$\begin{aligned} \mathbb{H}L(W, B, u) &= \mathbb{R}L(W_x, B_x, u_x) + \tau^2 \mathbb{R}L(W_y, 0, u_y) \\ &+ \tau \left( \mathbb{R}L(W_x, 0, u_y) + \mathbb{R}L(W_y, B_y, u_x) \right), \end{aligned}$$

где  $\mathbb{R}L(w, b, x)$  обозначает линейный слой с реальным значением, с  $w$ ,  $b$  и  $x$  для

весов, смещения и входных данных соответственно.

Операция среднего пула подразумевает вычисление среднего арифметического для каждого блока элементов. Это означает свертывание каждого  $n \times n$  блока элементов на его среднее значение. Это эквивалентно свёртке, когда шаг равен размеру ядра, где вес ядра равен действительным числам  $\frac{1}{n^2}$ . Так как ядро  $W = W_x + \tau W_y$  в этом случае полностью вещественное,  $W_y = 0$ , и формула конволюции может быть упрощена:

$$W * u = W(x + \tau y) = W_x * (x + \tau y) = W_x * x + \tau W_x * y.$$

Исходя из этого выражения, среднее пулирование эквивалентно двум вещественным операциям пулирования средних: каждая из них применяется независимо к каждому компоненту входных данных

$$u = x + \tau y:$$

$$\mathbb{H}AvgPool(u) = \mathbb{R}AvgPool(x) + \tau \mathbb{R}AvgPool(y)$$

Активационные функции используются для введения нелинейности в нейронные сети. Существует множество функций активации, основанных на вещественных числах, и большее их разнообразие основано на гиперкомплексных числах. Среди вещественных активаций существует семейство функций типа ReLU, которые помогают решить проблему затухающего градиента. Функции этого типа также используются в комплексной алгебре. Например, уже известно применение ReLU к реальным и мнимым частям отдельно [45]. В этом исследовании автор расширяет это определение и применяем его к другим алгебрам:

$$\mathbb{H}ReLU(u) = \mathbb{R}ReLU(x) + \tau \mathbb{R}ReLU(y).$$

## **Норма**

Традиционно в математической и физической литературе модуль комплексного

числа определяется как

$$\mathbb{C}|z| = \sqrt{zz^*} = \sqrt{(x + iy)(x - iy)} = \sqrt{x^2 + y^2}.$$

Однако обобщение этого метода на другие типы гиперкомплексных чисел не очень хорошо работает:

$$\mathbb{H}|z| = \sqrt{uu^*} = \sqrt{(x + \tau y)(x - \tau y)} = \begin{cases} |x|, & u \in \mathbb{D} \\ \sqrt{x^2 - y^2}, & u \in \mathbb{S} \end{cases}$$

Этот результат вряд ли применим к цели. Дуальная норма не зависит от дуальной части. В случае двойных чисел функция не определена для половины элементов. Поэтому нужно разработать другую формулу, которая распространяла бы стандартную норму комплексного числа на дуальные и двойные алгебры.

Чтобы определить выражение, обратимся к матричному представлению  $A = \begin{pmatrix} x & y \\ \sigma y & x \end{pmatrix}$  гиперкомплексного числа  $u = x + \tau y$ . Тогда автор связывает норму этой матрицы с нормой исходного дуального числа. Существует несколько способов определения матричной нормы. Во-первых, определим  $\mathbb{R}^{m \times n}$  как векторное пространство матриц с  $m$  строк и  $n$  столбцов записей в поле вещественных чисел  $\mathbb{R}$ . Используем норму матрицы, вызванную нормой вектора  $\|\cdot\|_2$  на  $\mathbb{R}^{2 \times 2}$  и вектор нормы  $\|\cdot\|_2$  на  $\mathbb{R}^{2 \times 1}$  и определим дуальную форму

$$\|u\|^2 = \sup\{\|At\|_2^2 : t \in \mathbb{R}^{2 \times 1}, \|t\|_2^2 = 1\}.$$

Для матрицы, которая соответствует гиперкомплексному числу  $u = x + \tau y$ , выражение нормы выглядит следующим образом:

$$\begin{aligned}
\|At\|_2^2 &= \left( \begin{pmatrix} x & y \\ \sigma y & x \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \right)^T \begin{pmatrix} x & y \\ \sigma y & x \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \\
&= (xt_1 + yt_2)^2 + (xt_2 + \sigma yt_1)^2 = \begin{cases} t_1^2 + t_2^2 = 1 \\ t_1 = \sin \varphi \\ t_2 = \cos \varphi \end{cases} \\
&= x^2 + y^2(\cos^2 \varphi + \sigma^2 \sin^2 \varphi) + xy \sin 2\varphi (1 + \sigma).
\end{aligned}$$

Для того, чтобы найти экстремум этой функции, приравняем производную к нулю и решаем для  $\varphi$ :

$$\begin{aligned}
f'(\varphi) &= y^2 \sin 2\varphi (\sigma^2 - 1) + 2xy \cos 2\varphi (1 + \sigma) = 0 \\
\tan 2\varphi &= -\frac{2x}{y(1 - \sigma)} \Rightarrow \begin{cases} \cos 2\varphi = \mp \frac{y(1 - \sigma)}{\sqrt{4x^2 + y^2(1 - \sigma)^2}} \\ \sin 2\varphi = \pm \frac{2x}{\sqrt{4x^2 + y^2(1 - \sigma)^2}} \end{cases}.
\end{aligned}$$

В конце концов, получаем максимум функции:

$$\|u\|^2 = x^2 + \frac{y^2(1 + \sigma^2)}{2} + |y|(1 + \sigma) \sqrt{x^2 + y^2 \left(\frac{1 - \sigma}{2}\right)^2}.$$

Из этого легко показать, что

$$\|z\| = \frac{|y|(1 + \sigma)}{2} + \sqrt{x^2 + y^2 \left(\frac{1 - \sigma}{2}\right)^2}.$$

Частные случаи комплексных, дуальных, двойных чисел ( $\sigma = -1, 0, 1$ ) приводят к следующей формуле для нормы гиперкомплексных чисел:

$$\mathbb{H}\|u\| = \mathbb{H}\|x + \tau y\| = \begin{cases} \sqrt{x^2 + y^2}, & u \in \mathbb{C} \\ \left| \frac{y}{2} \right| + \sqrt{x^2 + \left(\frac{y}{2}\right)^2}, & u \in \mathbb{D} \\ |x| + |y|, & u \in \mathbb{S} \end{cases}$$

### Гиперкомплексная пакетная нормализация.

Известная формула для пакетной нормализации [84]:

$$\hat{\mathbf{x}}^{(k)} = \frac{\mathbf{x}^{(k)} - \mathbb{E}[\mathbf{x}^{(k)}]}{\sqrt{V[\mathbf{x}^{(k)}]}}$$

где матрица ковариации  $V[\mathbf{x}^{(k)}]$  определяется как:

$$V[\mathbf{x}^{(k)}] = \begin{pmatrix} \text{Cov}(x_r^{(k)}, x_r^{(k)}) & \text{Cov}(x_r^{(k)}, x_i^{(k)}) \\ \text{Cov}(x_i^{(k)}, x_r^{(k)}) & \text{Cov}(x_i^{(k)}, x_i^{(k)}) \end{pmatrix}.$$

Подобно подходу с вещественными значениями, комплексная пакетная нормализация также имеет дополнительное линейное преобразование с двумя параметрами  $\hat{\gamma}^{(k)}$ ,  $\hat{\beta}^{(k)}$  с вещественными тренируемыми параметрами [45]:

$$\hat{\gamma}^{(k)} = \begin{pmatrix} \hat{\gamma}_{rr}^{(k)} & \hat{\gamma}_{ri}^{(k)} \\ \hat{\gamma}_{ir}^{(k)} & \hat{\gamma}_{ii}^{(k)} \end{pmatrix}, \quad \hat{\beta}^{(k)} = \begin{pmatrix} \hat{\beta}_r^{(k)} \\ \hat{\beta}_i^{(k)} \end{pmatrix}.$$

То есть, комплексная пакетная нормализация:

$$\mathbb{C}BN[\hat{\mathbf{x}}^{(k)}] = \hat{\gamma}^{(k)}\hat{\mathbf{x}}^{(k)} + \hat{\beta}^{(k)}.$$

Автор обобщает процесс пакетной нормализации для дуальных и двойных тензоров. Для достижения этой цели автор использует предложенную выше норму. Отметим некорректность использования той же процедуры, что и для комплексной пакетной нормализации (из [45]), для дуального входа по следующей причине:

$$\begin{aligned} \mu^{(k)} &= \mathbb{E}[\mathbf{x}^{(k)}] = \mu_r^{(k)} + \varepsilon\mu_d^{(k)}, \\ \Gamma^{(k)} &= \mathbb{E}\left[(\mathbf{x}^{(k)} - \mu^{(k)})^2\right] \\ &= \mathbb{E}\left[(x_r^{(k)} - \mu_r^{(k)})^2\right] \\ &\quad + \varepsilon\mathbb{E}\left[(x_r^{(k)} - \mu_r^{(k)})(x_d^{(k)} - \mu_d^{(k)})\right], \\ \mathbb{C}^{(k)} &= \mathbb{E}\left[(x^{(k)} - \mu^{(k)})\overline{(x^{(k)} - \mu^{(k)})}\right] = \mathbb{E}\left[(x_r^{(k)} - \mu_r^{(k)})^2\right]. \end{aligned}$$

Здесь видно, что не ковариация  $\Gamma$ , не псевдоковариация  $\mathbb{C}$  не зависят от  $\mathbb{E}\left[(x_d^{(k)} - \mu_d^{(k)})^2\right]$ . Эта проблема аналогична независимости нормы дуального числа

от ее дуальной части. Поэтому мы должны найти альтернативный способ обобщить пакетную нормализацию для всех типов гиперкомплексных алгебр второго порядка.

В этом исследовании автор представляет метод, основанный на концепции нормы для гиперкомплексных чисел, выведенной выше.

Во-первых, автор определяет среднее значение традиционным канальным способом, который одинаков для всех алгебр:

$$\mu^{(k)} = E[u^{(k)}] = \mu_x^{(k)} + \tau\mu_y^{(k)}.$$

Затем определяем дисперсию, специфичную для типа гиперкомплексной алгебры:

$$\text{Var}[u^{(k)}] = \frac{1}{m-1} \sum_{i=1}^m \|u_i^{(k)} - \mu^{(k)}\|^2,$$

где  $\|\cdot\|$  определена выше. Преобразуем входные данные как:

$$\hat{u}^{(k)} = \frac{u^{(k)} - \mu^{(k)}}{\sqrt{\text{Var}[u^{(k)}] + \delta}},$$

где  $\delta$  это небольшое число, необходимое для того, чтобы избежать деления на ноль.

Последним шагом является гиперкомплексное канальное масштабирование и смещение:

$$\text{HBN}[\hat{u}^{(k)}] = \hat{\gamma}^{(k)}\hat{u}^{(k)} + \hat{\beta}^{(k)},$$

где  $\hat{\gamma}^{(k)}$  и  $\hat{\beta}^{(k)}$  являются гиперкомплексными весами и смещением соответственно.

В частном случае комплексных чисел:

$$\begin{aligned} \text{CVar}[z] &= E[(z - \mu)(z - \mu)^*] \\ &= E[(x - \mu_x)^2] + E[(y - \mu_y)^2] = \Gamma(z). \end{aligned}$$

Итак, делим центрированный вход на  $\sqrt{\text{CVar}[z]}$ , что даёт ковариацию  $\Gamma(\hat{z}) = 1$ .

Заметим, что в общем случае псевдоковариация не равна нулю:

$$C(\hat{z}) = E[\hat{z}^2] = E[(\hat{x} + i\hat{y})^2] = E[\hat{x}^2 - \hat{y}^2] + 2iE[\hat{x}]E[\hat{y}] \\ \neq 0.$$

**Обратное распространение градиента функции потерь.** Важной частью обучения нейронной сети является градиентное вычисление. Вычисление градиента функции потерь опирается на алгоритм обратного распространения, который использует цепное правило.

В этой диссертации автор исследует проблему градиентного распространения в алгебре гиперкомплексных чисел. Классическое определение производной функции  $f(u) = f(x + \tau y) = v(x, y) + \tau w(x, y)$  гиперкомплексного элемента  $u = x + \tau y$ , где  $v$  и  $w$  есть вещественные функции:

$$f'(u) = \lim_{\Delta u \rightarrow 0} \frac{f(u + \Delta u) - f(u)}{\Delta u}.$$

Этот предел может существовать только в том случае, если он определен при  $\Delta u$  подходы к нулю вдоль реальной оси  $\Delta u = x$  или мнимой оси  $\Delta u = \tau \Delta y$ . В любом случае это должно дать один и тот же результат.

Уровняв эти частные случаи, получим обобщённые эквиваленты уравнений Коши-Римана:

$$\frac{\partial v}{\partial x} = \frac{\partial w}{\partial y} \quad \frac{\partial v}{\partial y} = \tau^2 \frac{\partial w}{\partial x} = \sigma \frac{\partial w}{\partial x}.$$

Функции, удовлетворяющие этим уравнениям, называются голоморфными. На практике это сильное ограничение, и большинство существующих операторов не удовлетворяют критериям Коши-Римана.

Чтобы преодолеть это, автор использует подход, изобретенный Виртингером для комплексных чисел. Он применяет подстановку переменных для того, чтобы переписать функцию комплексной переменной  $f(z)$  как голоморфные функции

двух переменных  $f(z, z^*)$ . Автор расширяет этот подход на все алгебры второго порядка:

$$x = \frac{u + u^*}{2} \quad y = \frac{u - u^*}{2\tau}.$$

Для комплексных и двойных чисел мы можем легко устранить  $\frac{1}{\tau}$  путем умножения числителя и знаменателя на  $\tau$  и с  $\tau^2 = \sigma \in \mathbb{R}$  в знаменателе.

Во-первых, определим, что именно должно быть рассчитано. В случае комплексных сетей, исследователи, как правило используют  $x_n$  и мнимую  $y_n$  части весов  $z_n$  как отдельные вещественные каналы, и обновляют их используя вещественную производную функции потерь:

$$\begin{aligned} x_{n+1} &= x_n - \alpha \frac{\partial L}{\partial x} \\ y_{n+1} &= y_n - \alpha \frac{\partial L}{\partial y} \end{aligned} \rightarrow z_{n+1} = z_n - \alpha \left( \frac{\partial L}{\partial x} + i \frac{\partial L}{\partial y} \right).$$

Здесь автор обобщит эти вычисления для всех гиперкомплексных чисел второго порядка, учитывая:

$$u_{n+1} = u_n - \alpha \left( \frac{\partial L}{\partial x} + \tau \frac{\partial L}{\partial y} \right).$$

Определим выражение внутри скобки как гиперкомплексный градиент и вычисляем его через градиент гиперкомплексного оператора  $f$  как:

$$\frac{\partial L}{\partial x} + \tau \frac{\partial L}{\partial y} = \frac{\partial L}{\partial f} \left( \frac{\partial f}{\partial x} + \tau \frac{\partial f}{\partial y} \right) + \left( \frac{\partial L}{\partial f} \right)^* \left( \frac{\partial f}{\partial x} - \tau \frac{\partial f}{\partial y} \right)^*.$$

Стоит отметить, что  $\frac{\partial L}{\partial x} + \tau \frac{\partial L}{\partial y}$  может быть выражено через  $u$  и  $u^*$ :

$$\frac{\partial f}{\partial x} + \tau \frac{\partial f}{\partial y} = \begin{cases} 2 \frac{\partial f}{\partial u^*} & u \in \mathbb{C} \\ 2 \frac{\partial f}{\partial u} & u \in \mathbb{S}. \\ \frac{\partial f}{\partial u} + \frac{\partial f}{\partial u^*} + \varepsilon^2 \left( \frac{\partial f}{\partial u} - \frac{\partial f}{\partial u^*} \right) & u \in \mathbb{D} \end{cases}$$

Автор реализовал эту формулу для дуальной алгебры. Этот подход показывает тот же результат, что и расчет двух производных по вещественной. Эксперименты показывают, что время тренировок тоже примерно одинаковое.

### **Конвертация в гиперкомплексные сети.**

Проблема заключается в преобразовании существующих архитектур нейронных сетей, основанных на вещественных числах, в сети, использующие гиперкомплексные значения. Это преобразование предназначено для представления нейронной сети в более общем виде и достижения более высокой точности за счет использования большего количества параметров. В качестве отправной точки должны быть выполнены следующие шаги:

1. Загрузить реально значимую модель с предварительно обученными весами. Точность модели, как ожидается, будет лучшей для выбранной архитектуры на определенном наборе данных.
2. Преобразовать модель в гиперкомплексную. Реальная часть весов инициализируется как исходные грузы, а мнимая часть заполняется нулями. Этот подход позволяет получить гиперкомплексную модель с той же метрикой точности, что и исходная модель.
3. Обучить полученную модель. Начиная с контрольной точки, которая дает относительно высокую точность, модель использует свою повышенную способность обобщения (в связи с добавлением мнимого канала для данных) получить более высокие результаты, чем исходная модель.

## Central Kernel Alignment (СКА) метрики

В этом разделе автор проводит сравнительный анализ признаков, порожденных вещественными и гиперкомплексными (дуальными, комплексными) моделями. Основная гипотеза состоит в том, что гиперкомплексные нейронные сети извлекают признаки, которые отличаются от признаков, сгенерированных вещественной моделью с той же архитектурой. Для проверки этой гипотезы введем несколько метрик подобия.

Автор проверил несколько стандартных методов, таких как:

- Корреляции (коэффициенты Кендалла  $\tau$  и Пирсона  $r$ ) между распределениями для задачи классификации, полученными из разных сетей;
- Анализ позиций векторов для вложений последнего слоя сетей, работающих в разных алгебрах;
- Разница между САМ, сгенерированными нейронными сетями для одних и тех же образцов.

Было показано, что эти методы статистически нестабильны и их трудно интерпретировать, поэтому используется метрика Central Kernel Alignment (СКА). Он представлен в виде матрицы, каждый элемент которой обозначает критерий независимости Гильберта-Шмидта (HSIC) для слоев двух разных сетей:

$$\text{HSIC}(K, L) = \frac{1}{b(b-3)} \left( \text{tr}(\tilde{K}\tilde{L}) + \frac{\sum_{i,j=1}^b K_{ij} \sum_{i,j=1}^b \tilde{L}_{ij}}{(b-1)(b-2)} - \frac{2}{b-2} \sum_{i,j=1}^b (K\tilde{L})_{ij} \right),$$

где  $\tilde{K}_{ij} = K_{ij} - \delta_{ij}K_{ij}$ ,  $b$  представляет собой размер пакета, и все тензоры  $\tilde{L}_{ij} = L_{ij} - \delta_{ij}L_{ij}$

сглаживаются из  $b \times c \times h \times w$  в 2D матрицы  $b \times c * h * w$ .

$СКА_{i,j}(X, Y) =$

$$\frac{\frac{1}{n} \sum_{k=1}^n \text{HSIC}(X_i^{(k)} X_i^{(k)T}, Y_j^{(k)} Y_j^{(k)T})}{\sqrt{\frac{1}{n} \sum_{k=1}^n \text{HSIC}(X_i^{(k)} X_i^{(k)T}, X_i^{(k)} X_i^{(k)T})} \sqrt{\frac{1}{n} \sum_{k=1}^n \text{HSIC}(Y_j^{(k)} Y_j^{(k)T}, Y_j^{(k)} Y_j^{(k)T})}}$$

Автор протестировал эту метрику на реальных моделях ResNet18, обученных на наборе данных CIFAR-100 с нуля для двух разных начальных значений. Для этого автор рассчитал метрику для пар слоев, взятых из разных моделей. Ожидалось, что скрытые представления этих моделей близки, потому что это одна и та же модель, обученная на одном и том же наборе данных.

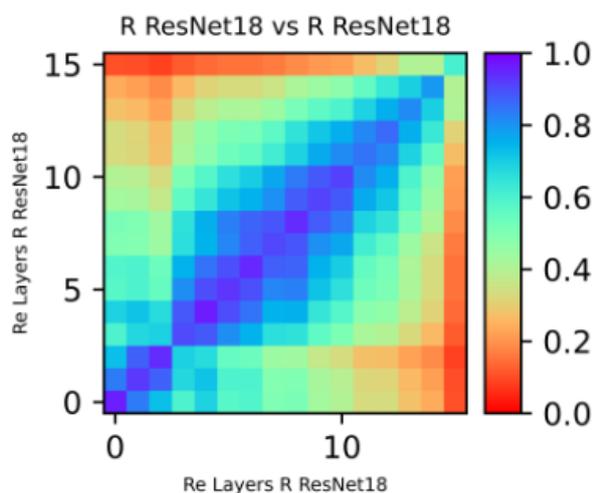


Рисунок 6. Значение метрики SKA для моделей ResNet18, обученных с разными начальными значениями.

Как и ожидалось, SKA близок к 1 для диагональных элементов, поскольку обе модели с одинаковым номером слоя должны иметь одинаковое представление. Исходная метрика SKA определена для наборов действительных чисел, поэтому автор расширяет этот подход для случаев гиперкомплексных алгебр. Автор

попробовал следующие методы:

- Матричное представление комплексных и дуальных чисел;
- Норма комплексных и дуальных чисел;
- Объединение дуальных/мнимых компонентов с реальными;
- Вычислять отдельно вещественные и гиперкомплексные части.

Предложенные методы дали схожие результаты, поэтому последний подход был выбран как более демонстративный.

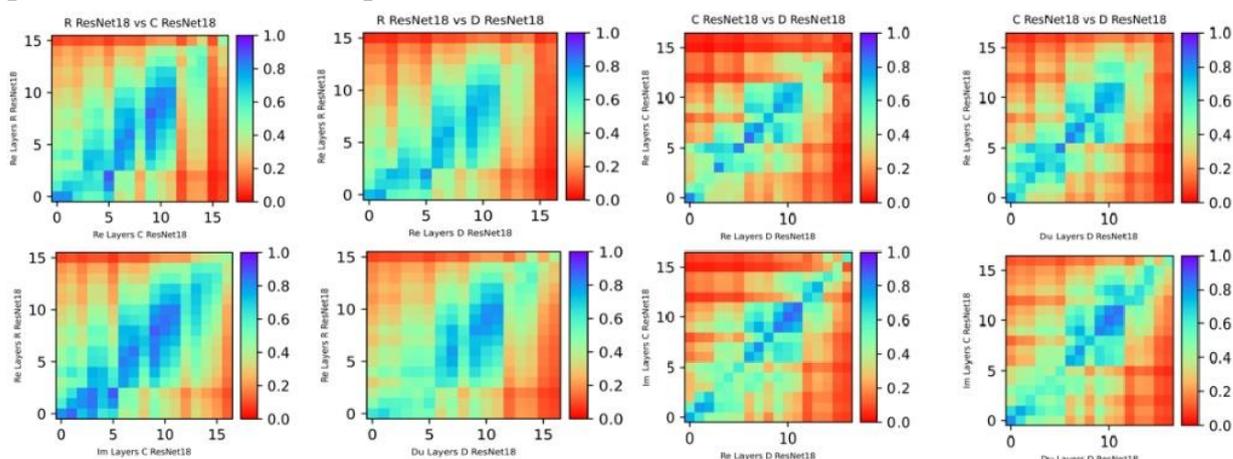


Рисунок 7. Значения метрики СКА, рассчитанные отдельно для вещественной и гиперкомплексной части для пары моделей ResNet18 из разных алгебр.

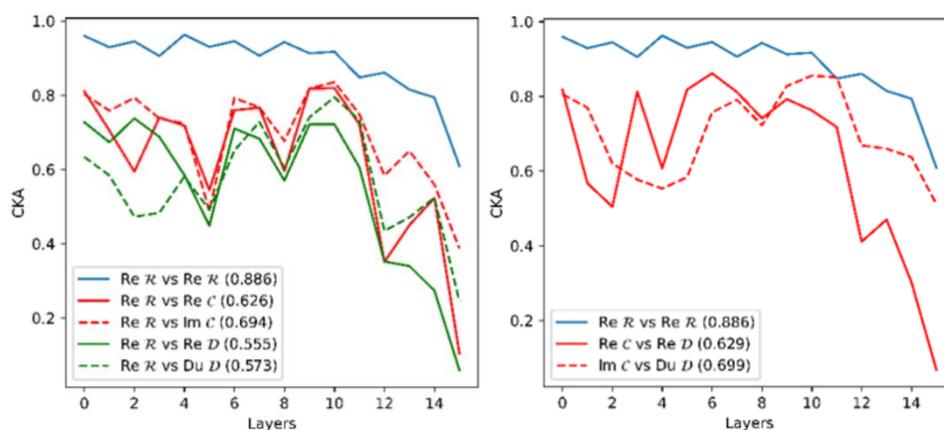


Рисунок 8. Диагональные значения метрик СКА. (слева) между выходной реальной НС и реальной/дуальной/комплексной частью гиперкомплексных НС

(справа) некоторые частные случаи.

Рисунок 8 показывает, что характеристики моделей, основанных на разных алгебрах, достаточно близки. Таким образом, делаем вывод о возможности переноса знаний для реальных нейронных сетей в гиперкомплексные и наоборот. Дальнейшие результаты дистилляции знаний также подтверждают это предположение.

В то же время средние значения СКА для  $\text{Re } \mathbb{R}$  и  $\text{Re } \mathbb{D}$  0.555 и для  $\text{Re } \mathbb{R}$  и  $\text{Du } \mathbb{D}$  0.573, таким образом, нейронные сети с дуальным значением имеют некоторые новые особенности (features), которые не представлены в моделях с действительными значениями. Эта метрика также показывает, что дуальные и комплексные значения достаточно близки,  $\text{Im } \mathbb{C}$  и  $\text{Du } \mathbb{D}$  0.699.

### **Передача знаний (ПЗ) из нейронных сетей с реальными значениями**

В предыдущем подразделе показано, что скрытое представление данных для реальной и гиперкомплексной моделей близко. Это побудило нас проверить возможность переноса знаний из вещественных моделей в гиперкомплексные.

Для этого использовались веса моделей ResNet с действительными значениями, предварительно обученные на наборе данных ImageNet из PyTorch, и инициализировали действительную часть весов гиперкомплексной модели этими значениями. Мнимая часть весов равна нулю.

Результаты проведенных экспериментов представлены в таблице 2. Видно, что использование предварительно обученных весов приводит к повышению точности моделей. Примечательным фактом является то, что даже после переноса знаний

гиперкомплексные (дуальные и комплексные) модели показывают лучшую точность, чем реальная.

Таблица 2. Результат (accuracy, %) обучения гиперкомплексных моделей ResNet18 с нагруженными весами, дата сет: CIFAR-100, задача: классификация.

	Без ПЗ, %	С ПЗ, %
Real	75.9	79.9
Dual	<b>78.3</b>	<b>82.1</b>
Complex	78.2	81.9

По результатам проведенных экспериментов делается вывод, что передача знаний способствует повышению сходимости дуальных моделей (рис. 9)

**Во второй главе** приведены реализации гиперкомплексных сетей, классические задачи компьютерного зрения, обнаружение гравитационных волн, задача по транскрипции музыки, а также дуальных голоморфных нейронных сетей.

### **Гиперкомплексная «игрушечная сеть» Toy Net**

Прежде чем перейти к моделям глубокого обучения, начнём с демонстрационной нейронной сети, предназначенной для прогнозирования значений зашумленных функций гиперкомплексного аргумента. Игрушечная сеть состоит из двух сверток и функции активации сигмовидного типа между ними. Эта архитектура остается одинаковой для всех алгебр, но для каждой алгебры используем свою реализацию операторов. Чтобы показать преимущество моделей на основе гиперкомплексных операторов, проведём сравнение их результатов с результатом вещественной модели той же архитектуры. Среднее стандартное

отклонение используется в качестве функции потерь. Эти четыре модели обучались на 1000 эпох с одинаковым количеством параметров для предсказания значений двух функций:  $Ai(u)$ - функция Эйри первого рода и  $J_3(u)$  - функция Бесселя первого рода третьего порядка. Также добавляем шумы с нормальным распределением к значениям функций обучающей выборки (и проверочной выборки).

Таблица 3. Toy Net – значение функции потерь.

Функция	$Ai(u)$			$J_3(u)$		
	$u \in \mathbb{C}$	$u \in \mathbb{D}$	$u \in \mathbb{S}$	$u \in \mathbb{C}$	$u \in \mathbb{D}$	$u \in \mathbb{S}$
Real	0.026	0.045	0.013	0.018	0.017	0.015
Complex	<b>0.008</b>	0.015	0.015	<b>0.009</b>	0.012	0.011
Dual	0.017	<b>0.009</b>	0.017	0.013	<b>0.009</b>	0.012
Double	0.023	0.050	<b>0.010</b>	0.017	0.018	<b>0.010</b>

Из таблицы 3 видно, что наименьшая среднеквадратическая ошибка (СКО) между гиперкомплексным значением функции и двухкомпонентным (гиперкомплексным) предсказанием моделей достигается нейронной сетью, принадлежащей к тому же типу гиперкомплексных чисел, что и исходная функция (и ее аргумент). Например, комплексная функция лучше всего предсказывается комплексной моделью и так далее. Таким образом, заключаем, что сети, основанные на гиперкомплексных числах, могут обучаться зависимостям или закономерностям, лежащим в соответствующей алгебре.

## Классические задачи Компьютерного зрения CV

Прежде чем перейти к решению задач классификации CV, необходимо уточнить процесс подготовки данных. Для преобразования вещественного изображения в гиперкомплексный формат воспользуемся более общим методом по сравнению с уже предложенным ранее (не нами), где авторы преобразуют вещественное изображение в сложную форму с помощью  $[R, G, B] \Rightarrow [R + iG, G + iB]$ , утверждая, что этот тип кодирования фиксирует корреляции каналов и изменение оттенка.

В этом исследовании автор обобщает эту идею и использует два типа предварительной обработки. Первый из них  $[R, G, B] \Rightarrow [R + \tau G, G + \tau B, B + \tau R]$ , которую автор называет сочетанием цветов (СЦ). Второй вариант — это линейное преобразование (ЛП) с обучаемыми параметрами, которые преобразуют  $[R, G, B]$  в шесть вещественных каналов, которые позже преобразуются в три гиперкомплексных канала. Этот тип предобработки выглядит более предпочтительным, поскольку позволяет модели определить наилучшее цветовое пространство для задачи. Кроме того, СЦ является частным случаем ЛП, где матрица преобразования фиксирована.

Берем архитектуру ResNet18 в качестве основы модели, которая обобщается на гиперкомплексные алгебры второго порядка путем замены вещественных операторов их аналогами в гиперкомплексных алгебрах. Используется стохастический градиентный спуск с импульсом 0,9 для оптимизации функций потерь с действительным знаком, рассматривая действительную и гиперкомплексную части как отдельные каналы с действительным знаком. Для преобразования гиперкомплексных признаков в вещественные, применяем

гиперкомплексную норму, связанную с соответствующей алгеброй. Перекрестная потеря энтропии между входом и целью используется в качестве критерия для этих проблем. Скорость обучения планируется по следующему правилу: 0,1 для первых 60 эпох, 0,02 для 61-120 эпох, 0,004 для 121-160 эпох, 0,0008 для 161-200 эпох. Результаты классификации изображений на CIFAR-10, CIFAR-100 и SVHN представлены в таблице 4.

Таблица 4. Точность (ассурагу, %) моделей (вещественная, дуальная, комплексная, двойная) для классификации на CIFAR-100, CIFAR-10, SVHN, с сочетанием цветов (СЦ) и линейным преобразованием (ЛП).

Датасет	CIFAR-100		CIFAR-10		SVHN	
Real	74.37		93.83		95.95	
Алгебра	Препроцессинг					
	ЛП	СЦ	ЛП	СЦ	ЛП	СЦ
Dual	76.30	76.12	94.27	94.32	96.35	96.23
Complex	77.12	76.65	94.56	94.45	96.35	96.04
Double	75.52	75.76	94.23	94.05	96.30	96.20

Из таблицы 4 видно, что все модели, основанные на числах второго порядка, достигают более высокой точности по сравнению с реальной. Кроме того, комплексная нейронная сеть показывает лучшие значения метрик по сравнению с другими гиперкомплексными моделями.

Из таблицы 5 видно, что переход от реальной модели к гиперкомплексной

значительно увеличивает вычислительную сложность. Кроме того, линейное преобразование требует больше времени, чем комбинирование цветов. Худшую производительность показывает сеть с комплексными значениями, но реализация представления Re-Im помогает сократить разрыв. Из таблицы 5 также видно, что использование групповых сверток приводит к замедлению модели. Диагональное представление двойных чисел приводит к сокращению времени вывода более чем в 1,5 раза. Тем не менее, двойные сети не являются оптимальной моделью, поскольку они показывают наихудшую точность среди всех гиперкомплексных моделей.

Таблица 5. Среднее время инференса ( $\mu\text{s}$ ) гиперкомплексных моделей ResNet-50 для CIFAR-100, размер батча 1.

Алгебра	Время инференса, $\mu\text{s}$	
	CPU	GPU
Real	20.07	4.04
Dual	85.74	12.81
Complex	114.82	15.85
Double	108.95	15.73

Из таблиц 4 и 5 делаем вывод, что нейронные сети на основе дуальных чисел представляют собой разумный баланс между увеличением вычислительной сложности и достижением большей точности.

## Обнаружение гравитационных волн

Эта часть посвящена решению задачи обнаружения сигналов с использованием гиперкомплексных сетей для набора данных G2Net [84]. Набор данных состоит из смоделированных зашумленных сигналов, похожих на гравитационные волны, зарегистрированных системой из трех наземных лазерных интерферометров: LIGO Hanford, LIGO Livingston и Virgo [85-87]. Как правило, гравитационные волны излучаются во время космических событий, таких как слияние черных дыр [85]. Набор данных G2Net содержит записи об эмулированных событиях того же характера.

Чтобы классифицировать исходный сигнал, предварительно обрабатываем данные для изображения, а затем пропускаем их через нейронные сети. Цель предварительной обработки — построить репрезентативную карту частот исходного сигнала. Алгоритм CQT считается эффективным для анализа гравитационных волн.

Таблица 6. Средние значения метрик моделей для обнаружения гравитационных волн.

Алгебра	Точность (accuracy), %	AUC ROC
Real	76.45	0.82
Complex	78.73	0.84
Dual	<b>79.24</b>	<b>0.85</b>
Double	77.41	0.84

Эта предварительная обработка передает временные ряды в частотный портрет (рисунок 13), который рассматривается как изображение на последующих шагах. Таким образом, изменения частотных характеристик сигнала в определенный момент отражаются как визуальные особенности, такие как специфическая форма и цвет, на полученном изображении. Для классификации изображений, полученных после CQT алгоритма [84], снова используем модель ResNet18, операторы которой меняются на соответствующие в различных алгебрах.

Модель оптимизирована с помощью алгоритма Стохастического градиентного спуска с импульсом 0,9 и весовым спадом  $5 \cdot 10^{-5}$  для L2 регуляризации. Регулярность повышается за счет добавления выпадающих данных со значением 0,2. Планировщик скорости обучения использует политику Cosine Annealing с  $T_{\max} = 6$ , и начальное значение скорости обучения установлено  $5 \cdot 10^{-2}$ . Продолжительность тренинга 30 эпох.

### **Задача транскрипции музыки**

В этой части показываем результаты автоматической транскрипции музыки. Эксперименты проводятся с набором данных MusicNet [88]. Чтобы повысить вычислительную эффективность, передискретизируем исходный сигнал с 44,1 кГц на 11 кГц, основываясь на алгоритме из [89]. Это позволяет нам снизить вычислительные затраты без каких-либо значительных потерь информации. Как в [88] записи '2303', '2382', '1819' используются в качестве тестового подмножества, а другие 327 файлов используются в качестве обучающего набора. Проводим все эксперименты с комплексным представлением частотного спектра. Для

вещественной модели рассматриваем вещественные и мнимые компоненты спектра как отдельные каналы.

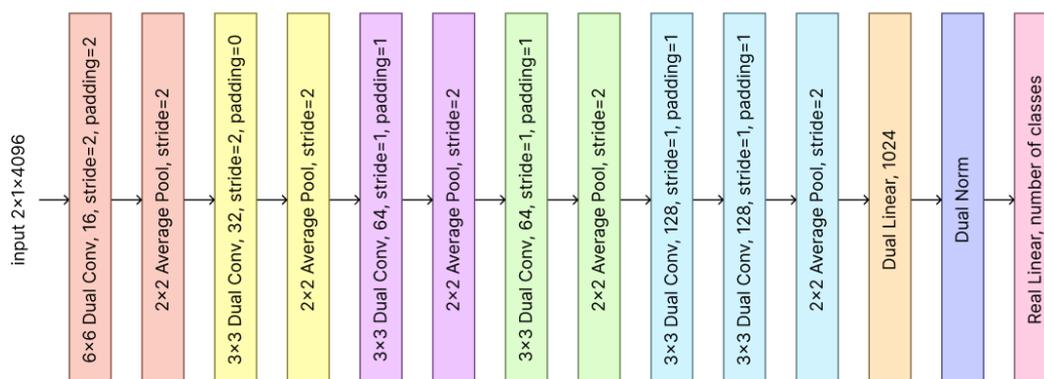


Рисунок 14. Дуальная архитектура DeepConvNet.

Структура одинакова в сложном случае вплоть до замены комплексных блоков на дуальные аналоги.

Также используем архитектуру DeepConvNet, разработанную в [45] (Рисунок 14). Сеть состоит из шести одномерных свёрточных слоев. Первый имеет фильтр размером шесть, а другие слои имеют размер ядра равным трем. За блоками свёртки следуют вещественный линейный слой с 2048 связями для реальной модели или комплексный/дуальный линейный слой с 1024 связями для комплексной/дуальной модели и функцией активации ReLU. Перед прохождением последнего слоя необходимо изменить представление данных с комплексной/дуальной формы на действительную. Чтобы сохранить всю информацию, объединяем реальные и мнимые/дуальные компоненты в один совместный канал. Наконец, применяем вещественный линейный слой с 84 связями и сигмоидной функцией активации. Количество единиц в последнем операторе соответствует количеству нот, присутствующих в наборе данных. Для вещественных, дуальных и комплексных моделей используем компонентную функцию активации ReLU, как это было

описано ранее. Во всех экспериментах используем окно ввода из 4096 отсчетов или его БПФ (что соответствует 8192-окну, используемому в базовой линии) и предсказывают ноты в центре окна. Все сети оптимизированы с помощью Adam [90]. Начинаем с скорости обучения на уровне  $10^{-3}$  для первых 10 эпох, а затем снижаем ее в 10 раз для каждой из эпох 10, 100, 120 и 150.

Комплексная сеть инициализируется с помощью унитарной схемы инициализации, соответствующей критерию He, как это было описано в [45]. Дуальные и вещественные модели инициализируются путем инициализации He в соответствии с методом, предложенным в [91], на основе равномерного распределения. Результаты резюмированы в таблице 7. Зависимость точности от отзыва показана на рис. 15.

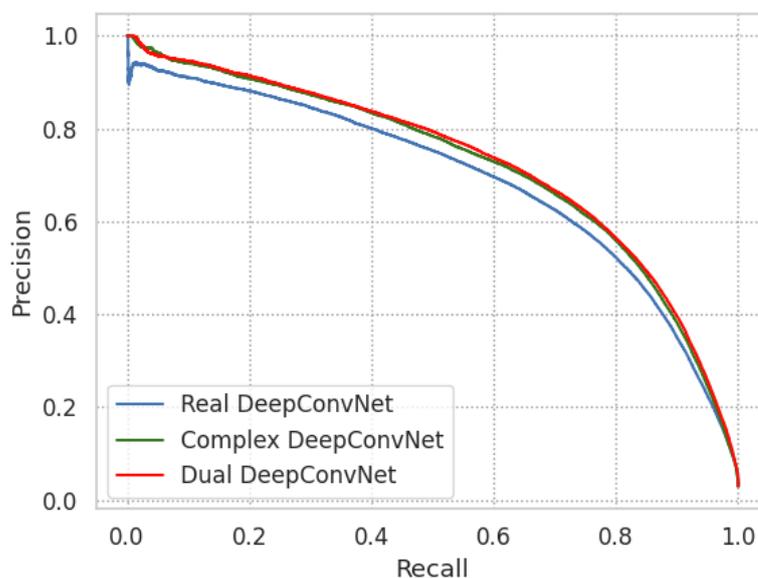


Рисунок 15. Кривая точности-отзыва для набора данных MusicNet.

Таблица 7. Сравнение основных метрик реальных и гиперкомплексных DeepConvNet моделей по задаче музыкальной транскрипции, датасет MusicNet.

Модель:	Точность (average precision), %	Количество параметров, МВ	Время инференса, $\mu$ s
DeepConvNet			
Real	68.9	34.2	<b>37</b>
Dual	<b>73.4</b>	34.2	51
Complex	73.2	34.2	64

Из таблицы 7 можно заметить, что наилучшая средняя точность достигается дуальной нейронной сетью. Кроме того, как будет показано, дуальная модель имеет  $\times 0.8$  времени инференса комплексной модели.

### Дуальные голоморфные нейронные сети

В этом разделе автор определяет дуальные голоморфные слои и показываем результаты моделей, основанных на них. «Голоморфные» означает, что эти слои удовлетворяют условиям Коши-Римана для дуальных значений функций.

#### 1) Условия Коши-Римана для функции дуальных переменных

Для того, чтобы комплексная функция была дифференцируемой, должно выполняться специальное уравнение, которое называется условиями Коши-Римана. Первоначально уравнения Коши — Римана были определенными критериями, необходимыми для комплексной функции  $f(x + iy) = u(x, y) + iv(x, y)$  быть голоморфными (сложно дифференцируемыми), где  $u$  и  $v$  являются вещественными функциями двух переменных. Эти уравнения накладывают ограничения на  $u(x, y)$  и  $v(x, y)$ :

$$\frac{\partial v}{\partial x} = \frac{\partial w}{\partial y}, \quad \frac{\partial v}{\partial y} = -\frac{\partial w}{\partial x}.$$

Существуют аналогичные условия для дуальной функции  $f(x + \varepsilon y) = u(x, y) + \varepsilon v(x, y)$  быть голоморфным (в смысле дуальных чисел):

$$\frac{\partial v}{\partial x} = \frac{\partial w}{\partial y}, \quad \frac{\partial v}{\partial y} = 0.$$

Используя разложение ряда Тейлора для дуального шага, можно показать, что вышеуказанные ограничения подразумевают, что голоморфная функция дуальной переменной разворачивается до следующего вида:

$$f(x + \varepsilon y) = f(x) + \varepsilon y f'(x).$$

В данном исследовании автор рассматривает дуальные функции и операторы как аналитическое продолжение вещественных. Таким образом, автор предполагает, что  $f(x)$  реален для любого реального  $x$ . Для ясности, это достаточное, но не обязательное условие для того, чтобы функция была голоморфной.

## 2) Дуальные голоморфные операторы

Общая формула дуальной свёртки для входа  $Z = X + \varepsilon Y$  и весов  $W = W_r + \varepsilon W_d$  со сдвигом  $b = b_r + \varepsilon b_d$  есть

$$Z * W + b = X * W_r + b_r + \varepsilon(Y * W_r + X * W_d + b_d).$$

Можно видеть, что общая формула дуальной свёртки в общем случае весовой матрицы  $W$  не удовлетворяет условию для дуальной функции быть голоморфным (в смысле дуальных чисел). Чтобы убедиться, что дуальная свёртка голоморфна, необходимо ввести ограничение  $Du(W) \equiv 0$ . Это условие основано на том факте, что для линейной функции  $f(x + \varepsilon y) = a_r x + b_r + \varepsilon(a_r y + a_i x + b_d)$  пределы его приращения, так как аргумент приближается к нулю вдоль вещественной оси или дуальной оси, равны тогда и только тогда, когда условие  $a_d = 0$  является верным.

Итак, получаем следующее уравнение для голоморфной дуальной свёртки:

$$Z * W + b = X * W_r + b_r + \varepsilon(Y * W_r + b_d).$$

Она имеет примерно в два раза меньше параметров, чем общая дуальная свёртка, или, другими словами, примерно такое же количество параметров, как вещественная.

Автор также определил несколько голоморфных дуальных функций активации:

$$\text{ReLU}(d) = \begin{cases} x + \varepsilon y, & x \geq 0 \\ 0 & x < 0 \end{cases} = \begin{cases} d, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$\sigma(d) = \sigma(x) + \varepsilon y \sigma(x)(1 - \sigma(x)), \text{ where } \sigma(x) = \frac{1}{1 + e^{-x}},$$

$$\tanh(d) = \tanh(x) + \varepsilon y (1 - \tanh(x)^2), \quad \text{where}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

### 3) Генерация дуального входного сигнала

Так же, как и комплексные, дуальные числа по существу являются парами вещественных значений. Исходя из этого сходства, предлагается использовать комплексный вход как в дуальных, так и в комплексных нейронных сетях. В этой работе используется два метода преобразования вещественных входных данных в сложный формат: преобразование Фурье для набора данных MusicNet и преобразование константы Q (CQT) для набора данных G2Net.

Автор также разрабатывает альтернативный вариант трансформации. Заметим, что  $Du(f(x + \varepsilon y))$  в основном определяется производной функции в той же точке, что и  $Re(f(x + \varepsilon y)) = f(x)$ . Основываясь на этом, предлагается преобразовать вещественные числа входных данных в дуальные числа следующим образом:

$$Input \rightarrow Input + \varepsilon(Input)',$$

где  $(Input)'$  как функция  $Input$ , который в некотором смысле автор называет производной от этого входа. Конкретное определение производной зависит от задачи. Например, если входной сигнал является временным рядом, то вполне естественно определить производную относительно времени как разность силы сигнала в соседних временных точках.

#### 4) Экспериментальные результаты

Чтобы проверить эту гипотезу, автор проводит несколько экспериментов с ранее использовавшимися наборами данных: G2Net, MusicNet и G2Net.

Автор также тестирует этот подход на наборе данных ImageNet.

Таблица 8. Точность (average precision) на наборе данных G2Net.

Model	Input	BN	Точность (average precision, %)
Real	$ CQT $	Real	76.5
Complex	$CQT$	Complex	78.7
Dual	$CQT$	Complex	73.5
Dual	$CQT$	Dual	<b>79.2</b>
Dual	$ CQT  + \varepsilon CQT '$	Dual	51.7
Holomorphic dual	$CQT$	Dual	77.0
Holomorphic dual	$ CQT  + \varepsilon CQT '$	Dual	77.6
Holomorphic dual	$ CQT  + \varepsilon CQT '$	Holomorphic dual	78.4

Таблица 9. Точность (average precision) на наборе данных MusicNet.

Model	Average precision, %
Real	68.9
Complex	<b>73.4</b>
Dual	73.2
Holomorphic dual	71.2

Таблица 10. Точность (average precision) на наборе данных ImageNet.

Model	Top-1	Top-5
Real	69.76	89.08
Dual	70.76	89.58
Holomorphic dual	<b>70.79</b>	<b>89.63</b>

Таблицы 8, 9 и 10 показывают, что, как и другие сети алгебр второго порядка, голоморфные дуальные модели показывают лучшие метрики, чем соответствующие модели с вещественными значениями, лишь немного отставая от дуальных моделей. На практике ускорение вывода зависит от архитектуры и составляет 10-25% по сравнению с дуальными моделями, что может стоить компромисса с некоторой точностью. Кроме того, голоморфные модели имеют примерно в два раза меньше параметров, чем дуальные модели той же архитектуры. Эти преимущества делают голоморфные дуальные сети жизнеспособным вариантом в случае аппаратных ограничений.

**В заключении** сформулированы основные результаты работы.

## Список литературы

- [1] M. Arjovsky, A. Shah, and Y. Bengio, “Unitary Evolution Recurrent Neural Networks,” vol. 48, 2015, [Online]. Available: <http://arxiv.org/abs/1511.06464>
- [2] T. Nitta, “The computational power of complex-valued neuron,” in Joint International Conference ICANN/ICONIP, 2003, pp. 993–1000.
- [3] R. Chakraborty, Y. Xing, and S. X. Yu, “SurReal: Complex-Valued Learning as Principled Transformations on a Scaling and Rotation Manifold,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 3, pp. 940–951, 2022, doi: 10.1109/TNNLS.2020.3030565.
- [4] U. Singhal, Y. Xing, and S. X. Yu, “Co-domain Symmetry for Complex-Valued Deep Learning,” Dec. 2021, Accessed: Mar. 14, 2022. [Online]. Available: <http://arxiv.org/abs/2112.01525>
- [5] J. Basse, L. Qian, and X. Li, “A Survey of Complex-Valued Neural Networks,” 2021, [Online]. Available: <http://arxiv.org/abs/2101.12249>
- [6] B. Widrow, J. M. McCool, and M. Ball, “The complex LMS algorithm,” *Proc. IEEE*, vol. 63, pp. 719–720, 1975.
- [7] A. Hirose and S. Yoshida, “Generalization Characteristics of Complex-Valued Feedforward Neural Networks in Relation to Signal Coherence,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, pp. 541–551, 2012.
- [8] T. Kim and T. Adalı, “Fully complex backpropagation for constant envelope signal processing,” *Neural Networks Signal Process. X. Proc. 2000 IEEE Signal Process. Soc. Work. (Cat. No.00TH8501)*, vol. 1, pp. 231–240 vol.1, 2000.
- [9] T. Kim and T. Adalı, “Fully Complex Multi-Layer Perceptron Network for

- Nonlinear Signal Processing,” *J. VLSI signal Process. Syst. signal, image video Technol.*, vol. 32, pp. 29–43, 2002.
- [10] S. Scardapane, S. Van Vaerenbergh, A. Hussain, and A. Uncini, “Complex-Valued Neural Networks With Nonparametric Activation Functions,” *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 4, pp. 140–150, 2018.
- [11] Y. Quan, D. Li, W. Zhenyong, C. Liu, and C. He, “Channel Estimation and Pilot Design for Uplink Sparse Code Multiple Access System based on Complex-Valued Sparse Autoencoder,” *IEEE Access*, 2019.
- [12] M. Catelani *et al.*, “MLMVNN for parameter fault detection in PWM DC-DC converters and its applications for buck DC-DC converter,” *2016 IEEE 16th Int. Conf. Environ. Electr. Eng.*, pp. 1–6, 2016.
- [13] A. Marseet and F. Sahin, “Application of complex-valued convolutional neural network for next generation wireless networks,” *2017 IEEE West. New York Image Signal Process. Work.*, pp. 1–5, 2017.
- [14] I. Cha and S. A. Kassam, “Channel Equalization Using Adaptive Complex Radial Basis Function Networks,” *IEEE J. Sel. Areas Commun.*, vol. 13, pp. 122–131, 1995.
- [15] S. Chen, S. McLaughlin, and B. Mulgrew, “Complex-valued radial basic function network, Part I: Network architecture and learning algorithms,” *Signal Process.*, vol. 35, pp. 19–31, 1994.
- [16] S. Chen, S. Mclaughlin, and B. Mulgrew, “Complex-valued radial basis function network, Part II: Application to digital communications channel equalisation,” *Signal Process.*, vol. 36, pp. 175–188, 1994.
- [17] D. Jianping, N. Sundararajan, and P. Saratchandran, “Communication channel

- equalization using complex-valued minimal radial basis function neural networks,” *IEEE Trans. neural networks*, vol. 13 3, pp. 687–696, 2002.
- [18] W. Gong, J. Liang, and D. Li, “Design of high-capacity auto-associative memories based on the analysis of complex-valued neural networks,” *2017 Int. Work. Complex Syst. Networks*, pp. 161–168, 2017.
- [19] A. Uncini, L. Vecci, P. Campolucci, and F. Piazza, “Complex-valued neural networks with adaptive spline activation function for digital-radio-links nonlinear equalization,” *IEEE Trans. Signal Process.*, vol. 47, pp. 505–514, 1999.
- [20] M. Scarpiniti, D. Vigliano, R. Parisi, and A. Uncini, “Generalized splitting functions for blind separation of complex signals,” *Neurocomputing*, vol. 71, pp. 2245–2270, 2008.
- [21] R. Huang and M.-S. Chen, “Adaptive equalization using complex-valued multilayered neural network based on the extended Kalman filter,” *WCC 2000 - ICSP 2000. 2000 5th Int. Conf. Signal Process. Proceedings. 16th World Comput. Congr. 2000*, vol. 1, pp. 519–524 vol.1, 2000.
- [22] M. Solazzi, A. Uncini, E. D. Di Claudio, and R. Parisi, “Complex discriminative learning Bayesian neural equalizer,” *ISCAS’99. Proc. 1999 IEEE Int. Symp. Circuits Syst. VLSI (Cat. No.99CH36349)*, vol. 5, pp. 343–346 vol.5, 1999.
- [23] N. Benvenuto, M. Marchesi, F. Piazza, and A. Uncini, “Non linear satellite radio links equalized using blind neural networks,” *[Proceedings] ICASSP 91 1991 Int. Conf. Acoust. Speech, Signal Process.*, pp. 1521–1524 vol.3, 1991.
- [24] A. B. Suksmono and A. Hirose, “Adaptive Beamforming by Using Complex-Valued Multi Layer Perceptron,” 2003.

- [25] A. Hirose and M. Kiuchi, "Coherent optical associative memory system that processes complex-amplitude information," *IEEE Photonics Technol. Lett.*, vol. 12, pp. 564–566, 2000.
- [26] Y. Chistyakov, E. Kholodova, A. Minin, H.-G. Zimmermann, and A. Knoll, "Modeling of electric power transformer using complex-valued neural networks," 2011.
- [27] A. Minin, Y. Chistyakov, E. Kholodova, H.-G. Zimmermann, and A. Knoll, "Complex Valued Open Recurrent Neural Network for Power Transformer Modeling," *J. Appl. Math. & informatics*, vol. 6, pp. 41–48, 2012.
- [28] J. Zhang and Y. Wu, "A New Method for Automatic Sleep Stage Classification," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, pp. 1097–1110, 2017.
- [29] S. Liu *et al.*, "A Multilevel Artificial Neural Network Nonlinear Equalizer for Millimeter-Wave Mobile Fronthaul Systems," *J. Light. Technol.*, vol. 35, pp. 4406–4417, 2017.
- [30] M. Peker, B. Şen, and D. Delen, "A Novel Method for Automated Diagnosis of Epilepsy Using Complex-Valued Classifiers," *IEEE J. Biomed. Heal. Informatics*, vol. 20, pp. 108–118, 2016.
- [31] S. Hu and A. Hirose, "Proposal of Millimeter-Wave Adaptive Glucose-Concentration Estimation System Using Complex-Valued Neural Networks," *IGARSS 2018 - 2018 IEEE Int. Geosci. Remote Sens. Symp.*, pp. 4074–4077, 2018.
- [32] S. Hu, S. Nagae, and A. Hirose, "Millimeter-Wave Adaptive Glucose Concentration Estimation With Complex-Valued Neural Networks," *IEEE Trans. Biomed. Eng.*, vol. 66, pp. 2065–2071, 2017.

- [33] Y. Suzuki and M. Kobayashi, “Complex-valued bidirectional auto-associative memory,” *2013 Int. Jt. Conf. Neural Networks*, pp. 1–7, 2013.
- [34] T. Ding and A. Hirose, “Fading Channel Prediction Based on Combination of Complex-Valued Neural Networks and Chirp Z-Transform,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 25, pp. 1686–1695, 2014.
- [35] C.-A. Popa, “Deep Hybrid Real-Complex-Valued Convolutional Neural Networks for Image Classification,” *2018 Int. Jt. Conf. Neural Networks*, pp. 1–6, 2018.
- [36] M. Matlacz and G. Sarwas, “Crowd counting using complex convolutional neural network,” *2018 Signal Process. Algorithms, Archit. Arrange. Appl.*, pp. 88–92, 2018.
- [37] I. N. Aizenberg and A. Gonzalez, “Image Recognition using MLMVN and Frequency Domain Features,” *2018 Int. Jt. Conf. Neural Networks*, pp. 1–8, 2018.
- [38] P. Virtue, S. X. Yu, and M. Lustig, “Better than real: Complex-valued neural nets for MRI fingerprinting,” *2017 IEEE Int. Conf. Image Process.*, pp. 3953–3957, 2017.
- [39] E. Aizenberg and I. N. Aizenberg, “Batch linear least squares-based learning algorithm for MLMVN with soft margins,” *2014 IEEE Symp. Comput. Intell. Data Min.*, pp. 48–55, 2014.
- [40] I. N. Aizenberg, S. Alexander, and J. Jackson, “Recognition of Blurred Images Using Multilayer Neural Network Based on Multi-valued Neurons,” *2011 41st IEEE Int. Symp. Mult. Log.*, pp. 282–287, 2011.
- [41] I. N. Aizenberg, D. Paliy, J. M. Zurada, and J. Astola, “Blur Identification by Multilayer Neural Network Based on Multivalued Neurons,” *IEEE Trans. Neural*

- Networks*, vol. 19, pp. 883–898, 2008.
- [42] I. N. Aizenberg, N. N. Aizenberg, C. Butakoff, and E. Farberov, “Image recognition on the neural network based on multi-valued neurons,” *Proc. 15th Int. Conf. Pattern Recognition. ICPR-2000*, vol. 2, pp. 989–992 vol.2, 2000.
- [43] I. N. Aizenberg and C. Butakoff, “Image processing using cellular neural networks based on multi-valued and universal binary neurons,” *J. VLSI signal Process. Syst. signal, image video Technol.*, vol. 32, pp. 169–188, 2000.
- [44] R. S. Zemel, C. K. I. Williams, and M. C. Mozer, “Lending direction to neural networks,” *Neural Networks*, vol. 8, pp. 503–512, 1995.
- [45] C. Trabelsi, O. Bilaniuk, Dmitriy Serdyuk, Sandeep Subramanian, J. F. Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, C. Pal, “Deep complex networks,” International Conference on Learning Representations, Poster, 2018.
- [46] S. Amilia, M. D. Sulistiyo, and R. N. Dayawati, “Face image-based gender recognition using complex-valued neural network,” *2015 3rd Int. Conf. Inf. Commun. Technol.*, pp. 201–206, 2015.
- [47] M. Miyauchi, M. Seki, A. Watanabe, and A. Miyauchi, “Interpretation of Optical Flow Through Complex Neural Network,” 1993.
- [48] M. Miyauchi, M. Seki, A. Watanabe, and A. Miyauchi, “Interpretation of optical flow through neural network learning,” *[Proceedings] Singapore ICCS/ISITA '92*, pp. 1247–1251 vol.3, 1992.
- [49] A. Hirose, T. Higo, and K. Tanizawa, “Holographic Three-Dimensional Movie Generation with Frame Interpolation Using Coherent Neural Networks,” *2006 IEEE Int. Jt. Conf. Neural Netw. Proc.*, pp. 492–497, 2006.

- [50] A. Hirose, T. Higo, and K. Tanizawa, "Efficient generation of holographic movies with frame interpolation using a coherent neural network," *IEICE Electron. Express*, vol. 3, pp. 417–423, 2006.
- [51] Y. Liu, H. Huang, and T. Huang, "Gain parameters based complex-valued backpropagation algorithm for learning and recognizing hand gestures," *2014 Int. Jt. Conf. Neural Networks*, pp. 2162–2166, 2014.
- [52] R. F. Olanrewaju, O. O. Khalifa, A. H. Abdulla, and A. M. Z. M. Khedher, "Detection of alterations in watermarked medical images using Fast Fourier Transform and Complex-Valued Neural Network," *2011 4th Int. Conf. Mechatronics*, pp. 1–6, 2011.
- [53] R. Hata and K. Murase, "Multi-valued autoencoders for multi-valued neural networks," *2016 Int. Jt. Conf. Neural Networks*, pp. 4412–4417, 2016.
- [54] Y. Kominami, H. Ogawa, and K. Murase, "Convolutional neural networks with multi-valued neurons," *2017 Int. Jt. Conf. Neural Networks*, pp. 2673–2678, 2017.
- [55] C.-A. Popa, "Complex-Valued Deep Boltzmann Machines," *2018 Int. Jt. Conf. Neural Networks*, pp. 1–8, 2018.
- [56] L. Li, L. Wang, F. L. Teixeira, C. Liu, A. Nehorai, and T. jun Cui, "DeepNIS: Deep Neural Network for Nonlinear Electromagnetic Inverse Scattering," *IEEE Trans. Antennas Propag.*, vol. 67, pp. 1819–1825, 2018.
- [57] S. Gu and L. Ding, "A Complex-Valued VGG Network Based Deep Learning Algorithm for Image Recognition," *2018 Ninth Int. Conf. Intell. Control Inf. Process.*, pp. 340–343, 2018.
- [58] D. Hayakawa, T. Masuko, and H. Fujimura, "Applying Complex-Valued Neural

- Networks to Acoustic Modeling for Speech Recognition,” *2018 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, pp. 1725–1731, 2018.
- [59] M. Kataoka, M. Kinouchi, and M. Hagiwara, “Music information retrieval system using complex-valued recurrent neural networks,” *SMC’98 Conf. Proceedings. 1998 IEEE Int. Conf. Syst. Man, Cybern. (Cat. No.98CH36218)*, vol. 5, pp. 4290–4295 vol.5, 1998.
- [60] M. Kinouchi and M. Hagiwara, “Memorization of melodies by complex-valued recurrent network,” *Proc. Int. Conf. Neural Networks*, vol. 2, pp. 1324–1328 vol.2, 1996.
- [61] A. Y. H. Al-Nuaimi, M. F. Amin, and K. Murase, “Enhancing MP3 encoding by utilizing a predictive Complex-Valued Neural Network,” *2012 Int. Jt. Conf. Neural Networks*, pp. 1–6, 2012.
- [62] Y.-S. Lee, C.-Y. Wang, S.-F. Wang, J.-C. Wang, and C.-H. Wu, “Fully complex deep neural network for phase-incorporating monaural source separation,” *2017 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 281–285, 2017.
- [63] C. S. Tay, K. Tanizawa, and A. Hirose, “Error Reduction in Holographic Movies Using a Hybrid Learning Method in Coherent Neural Networks,” *ICANN, Lecture Notes in Computer Science*, vol 4668, Springer, 2007.
- [64] J. Gao, B. Deng, Y. Qin, H. Wang, and X. Li, “Enhanced Radar Imaging Using a Complex-Valued Convolutional Neural Network,” *IEEE Geosci. Remote Sens. Lett.*, vol. 16, pp. 35–39, 2017.
- [65] M. Wilmanski, C. Kreucher, and A. O. Hero, “Complex input convolutional neural networks for wide angle SAR ATR,” *2016 IEEE Glob. Conf. Signal Inf. Process.*,

- pp. 1037–1041, 2016.
- [66] I. N. Aizenberg and C. Moraga, “Multilayer Feedforward Neural Network Based on Multi-valued Neurons (MLMVN) and a Backpropagation Learning Algorithm,” *Soft Comput.*, vol. 11, pp. 169–183, 2006.
- [67] K. Oyama and A. Hirose, “Adaptive phase-singular-unit restoration with entire-spectrum-processing complex-valued neural networks in interferometric SAR,” *Electron. Lett.*, vol. 54, pp. 43–45, 2018.
- [68] X. Yao, X. Shi, and F. Zhou, “Complex-Value Convolutional Neural Network for Classification of Human Activities,” *2019 6th Asia-Pacific Conf. Synth. Aperture Radar*, pp. 1–6, 2019.
- [69] X. Yao, X. Shi, and F. Zhou, “Human Activities Classification Based on Complex-Value Convolutional Neural Network,” *IEEE Sens. J.*, vol. 20, pp. 7169–7180, 2020.
- [70] T. Dong and T. Huang, “Neural Cryptography Based on Complex-Valued Neural Network,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 31, pp. 4999–5004, 2019.
- [71] S. Jankowski, A. Lozowski, and J. M. Zurada, “Complex-valued multistate neural associative memory,” *IEEE Trans. neural networks*, vol. 7 6, pp. 1491–1496, 1996.
- [72] T. Miyajima, F. Baisho, K. Yamanaka, K. Nakamura, and M. Agu, “A Phasor Model with Resting States,” *IEICE Trans. Inf. Syst.*, vol. 83, pp. 299–301, 2000.
- [73] H. H. Cevik, Y. E. Acar, and M. Çunkaş, “Day Ahead Wind Power Forecasting Using Complex Valued Neural Network,” *2018 Int. Conf. Smart Energy Syst. Technol.*, pp. 1–6, 2018.
- [74] T. Kitajima and T. Yasuno, “Output prediction of wind power generation system using complex-valued neural network,” *Proc. SICE Annu. Conf. 2010*, pp. 3610–

- 3613, 2010.
- [75] D. P. Mandic, S. Javidi, S. L. Goh, A. Kuh, and K. Aihara, “Complex-valued prediction of wind profile using augmented complex statistics,” *Renew. Energy*, vol. 34, pp. 196–201, 2009.
- [76] Y. Maeda, T. Fujiwara, and H. Ito, “Robot control using high dimensional neural networks,” *2014 Proc. SICE Annu. Conf.*, pp. 738–743, 2014.
- [77] I. Nishikawa, K. Sakakibara, T. Iritani, and Y. Kuroe, “2 types of complex-valued Hopfield networks and the application to a traffic signal control,” *Proceedings. 2005 IEEE Int. Jt. Conf. Neural Networks, 2005.*, vol. 2, pp. 782–787 vol. 2, 2005.
- [78] I. Nishikawa, T. Iritani, and K. Sakakibara, “Improvements of the Traffic Signal Control by Complex-Valued Hopfield Networks,” *2006 IEEE Int. Jt. Conf. Neural Netw. Proc.*, pp. 459–464, 2006.
- [79] J. Hu, Z. Li, Z. Hu, D. Yao, and J. Yu, “Spam Detection with Complex-Valued Neural Network Using Behavior-Based Characteristics,” *2008 Second Int. Conf. Genet. Evol. Comput.*, pp. 166–169, 2008.
- [80] F. M. Dimentberg, *The screw calculus and its applications in mechanics*. WP-AFB, Ohio : Foreign Technology Division, 1968.
- [81] A. Güneş Baydin, B. A. Pearlmutter, A. Andreyevich Radul, and J. Mark Siskind, “Automatic differentiation in machine learning: A survey,” *J. Mach. Learn. Res.*, vol. 18, pp. 1–43, 2018.
- [82] R. Kiran and K. Khandelwal, “Automatic implementation of finite strain anisotropic hyperelastic models using hyper-dual numbers,” *Comput. Mech.*, vol. 55, no. 1, pp. 229–248, 2015, doi: 10.1007/s00466-014-1094-1.

- [83] Y. Okawa and T. Nitta, “Learning Properties of Feedforward Neural Networks Using Dual Numbers,” in *Proceedings, APSIPA Annual Summit and Conference*, 2021, pp. 187–192.
- [84] The G2Net Dataset: [site]: URL: <https://www.kaggle.com/competitions/g2net-gravitational-wave-detection/data>.
- [85] B. P. Abbott and others, “Observation of Gravitational Waves from a Binary Black Hole Merger,” *Phys. Rev. Lett.*, vol. 116, no. 6, 2016, doi: 10.1103/PhysRevLett.116.061102.
- [86] R. M. Shannon and others, “Gravitational waves from binary supermassive black holes missing in pulsar observations,” *Science (80-. )*, vol. 349, no. 6255, pp. 1522–1525, 2015, doi: 10.1126/science.aab1910.
- [87] K. S. Thorne, “Gravitational Waves from Compact Bodies.” arXiv, 1995. doi: 10.48550/ARXIV.GR-QC/9506084.
- [88] J. Thickstun, Z. Harchaoui, and S. M. Kakade, “Learning features of music from scratch,” *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, pp. 1–14, 2017.
- [89] J. O. Smith, “Digital Audio Resampling.” 2002. [Online]. Available: <https://ccrma.stanford.edu/~jos/resample/>
- [90] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [91] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, Dec. 2016, doi: 10.1109/CVPR.2016.90.