



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

«Дальневосточный федеральный университет»

ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ

Департамент автоматизации и робототехники

Научно-исследовательская работа

на тему

**«Система распознавания и слежения за подводными биологическими
объектами с использованием АНПА»**

Выполнил студент группы

Б3118-15.03.06мхрт

_____ Феклин Н.П.

Проверил доцент

_____ Кацурин А.А.

(оценка)

г. Владивосток

2021

Оглавление

Вводная часть	3
1.1. Цель работы.....	3
1.2. Поставленные задачи.....	3
2. Теория нейронных сетей	4
2.1. Основные понятия	4
2.2. Анализ качества обучения нейронной сети	6
3. Исследование готовых решений.....	9
3.1. Зависимость среднего модуля веса синапса от размера нейросети	9
3.2. EagleEye	10
3.3. Поиск архитектуры нейронной сети.....	11
Вывод.....	15
Список литературы	16

Вводная часть

При определении архитектуры нейронной сети, необходимой для реализации системы распознавания и слежения за подводными биологическими объектами с использованием АНПА, во главе вопроса возникла проблема вычислительных мощностей. Так как подводный аппарат не имеет на борту излишков памяти, любые системы обнаружения должны максимально эффективно использовать предоставленные возможности. В случае искусственной нейронной сети её эффективность сильно зависит от ее размера и структуры, то есть от числа скрытых слоев и количества нейронов в них.

НС является сложным и трудно прогнозируемым средством решения задач. Зачастую выбор ее оптимального размера происходит эмпирическим способом, так как невозможно понять точную зависимость качеств обучения от количества нейронов. Но есть методы, позволяющие избежать полной случайности выбора.

1.1. Цель работы

Определить способы формирования размера и архитектуры нейронных сетей.

1.2. Поставленные задачи

1. Разобрать основные понятия нейронных сетей
2. Разобрать понятия качества обучения нейронных сетей
3. Рассмотреть методы подбора количества нейронов в скрытых слоях
4. Изучить вопрос автоматической оптимизации количества нейронов в скрытых слоях
5. Рассмотреть методы подбора подходящей архитектуры для НС

2. Теория нейронных сетей

2.1. Основные понятия

Тема нейронных сетей является очень обширной, поэтому для исследования готовых решений приведены только основные определения, понятия и правила. Для изучения данного вопроса были использованы следующие источники: [1][2][3].

Согласно [2]: Нейронная сеть - компьютерная программа, которая работает по принципу естественной нейронной сети в мозгу. Задача таких искусственных нейронных сетей - выполнять такие когнитивные функции, как решение проблем и машинное обучение. Основная привлекательность нейронных сетей - их способность имитировать умение мозга распознавать образы.

Приведенное определение указывает на сопоставление способности мышления человеческого мозга и вычислительных возможностей электронных устройств. Для такого подхода необходимо, чтобы информация о предметной области не записывалась явно в программу, а распределялась по всей сети в виде связей между элементами, выполняющими функцию человеческих нейронов. Для более ясного понимания данного определения необходимо четкое представление понятия машинного обучения, которое очень подробно разобрано в [1].

Как сказано в [3]: Принципиальным моментом появления нейрокомпьютеров является отказ от булевского логического базиса на уровне элементов вычислительных машин с переходом на базис пороговой логики, который в простейшем варианте моделирует функции нервной клетки. При этом изменяется не только элементарная база, но и архитектура ЭВМ.

Согласно этому тезису, для построения отдельных узлов и блоков ЭВМ необходимо использование линейных пороговых элементов (нейронов).

Нейрон принимает на вход n логических значений, выполняет над ними математические преобразования и возвращает на выход уже одно значение. То есть нейрон осуществляет логическое преобразование входных сигналов в выходной. Графическое представление нейрона изображено на рисунке 1.

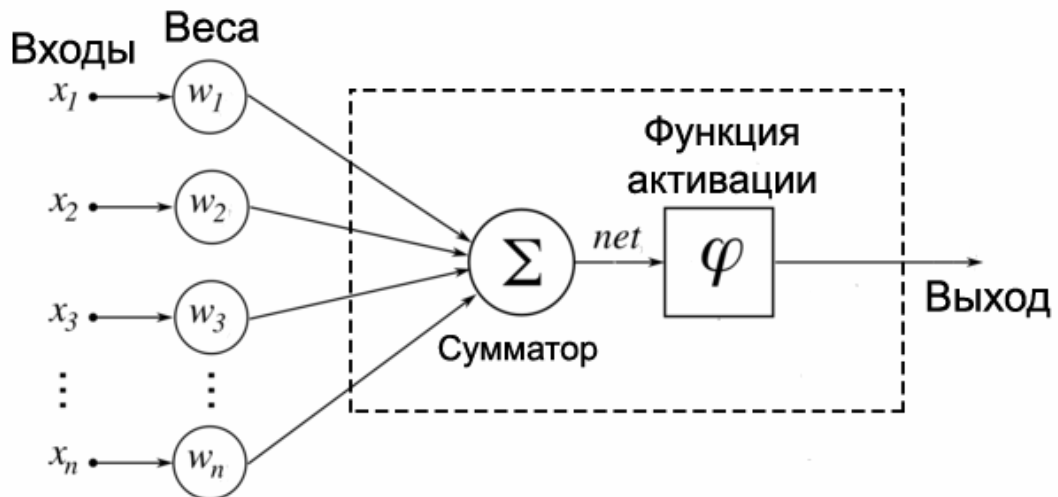


Рисунок 1– Линейный пороговый элемент (нейрон)

Рисунок 1 представляет собой графическое описание формулы:

$$y = \varphi \left(\sum_{i=0}^N \omega_i x_i \right) \quad (1)$$

$$\varphi(g) = \text{sign}(g) = \begin{cases} 0, & g < 0 \\ 1, & g \geq 0, \end{cases} \quad (2)$$

$$\varphi(g) = \begin{cases} 0, & g < 0 \\ g, & g \geq 0 \end{cases} \quad (3)$$

Функция (1) является простейшей интерпретацией передаточной функции нейрона, где y , x_i – значения выходного и входных нейронов соответственно ($x_i, y \in \{0,1\}$), ω_i – весовые коэффициенты, а $\varphi(g)$ – функция активации, она может иметь вид (2) или (3).

Нейронная сеть же будет являться совокупностью нескольких нейронов. Пример изображен на рисунке 2.

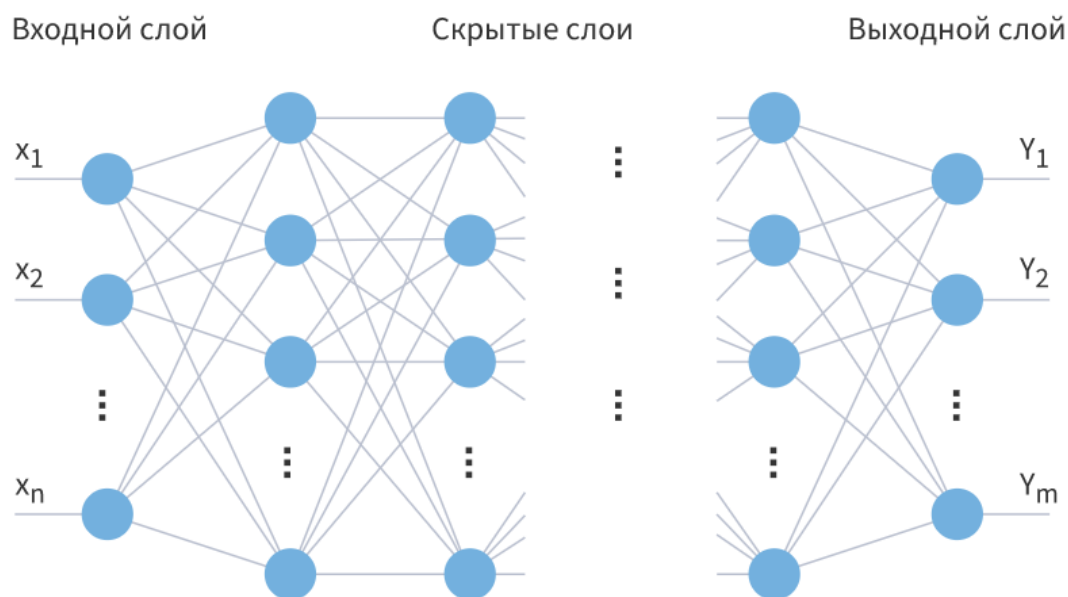


Рисунок 2 – Пример нейронной сети

Скрытый слой – это любой слой между входным (первым) слоем сети и выходным (последним).

2.2. Анализ качества обучения нейронной сети

При описании способности искусственной нейронной сети определить объект на изображении выделяются две величины.

Первая – это способность определить объект на изображении, которое она уже видела и, в теории, подстраивала нейроны под него. Благодаря этой величине можно судить о необходимости продолжать обучение. Если же эта величина перестала изменяться после какого-то момента и не достигла желаемых значений, то можно утверждать, что данная модель не удовлетворяет условиям решения поставленной задачи, и как минимум нужно изменять количество нейронов. Данная величина не характеризует качество нейронной сети, так как оно заключается в способности нахождения объектов на неизвестных изображениях.

Способность искусственной нейронной сети классифицировать объекты на изображениях, которые не участвовали в обучающей выборке,

называется обобщающая способность. Стремление этой величины к 100% и будет означать максимально правильно решенную задачу.

Качество работы сети зависит не только от ее основного параметра (вес нейронов), но и от гиперпараметров сети:

- Количество слоев и нейронов
- Количество эпох обучения
- Размер выборки обучения
- Размер параметра скорости обучения
- Архитектура НС

Проблема выбора гиперпараметров заключается во возникновении переобучения, вследствие которого обобщающая способность начнет уменьшаться.

В теории автоматизации проблема переобучения описывается графиками изображенными на рисунке 3

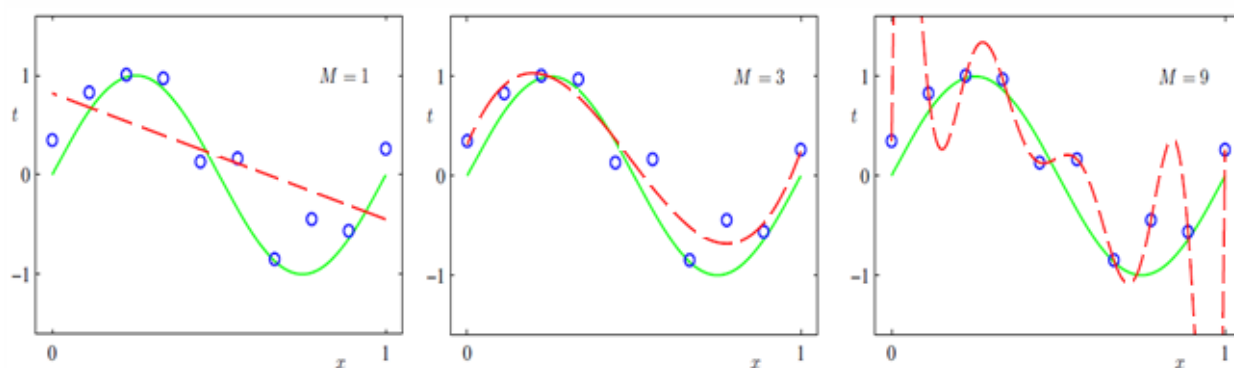


Рисунок 3 – Графическое представление проблемы переобучения

Как видно на рисунке 3, в первом случае система обучена недостаточно поэтому представление сильно искажается, но в случае 3, когда график старается точно интерпретировать полученные значения, появляется еще большее отклонения от желаемых результатов.

Проблема переобучения НС – это проблема, возникающая при обучении глубоких нейронных сетей в следствии долго обучения с использованием

неизменной обучающей выборки. Она характеризуется тем, что нейроны сети со временем адаптируются к определенным особенностям и перестают классифицировать объекты, не принимавшие участие в обучении. Объясним это явление на практическом примере обучения способности опознавать рыб.

Если обучающая выборка будет состоять из конкретных рыб, изображенных на рисунке 4 и обучение будет проходить излишнее количество итераций, то рыба, изображенная на рисунке 5, не будет классифицирована как рыба. Причина этого заключается в том, что рыбы, имеющей более треугольную форму в обучающей выборке, не было. При отсутствии переобучения обязательно наличие достаточного количества похожих участков картинок для опознания.



Рисунок 4 – Обучающая выборка



Рисунок 5 – Тестовый пример

В следствие описанных выше проблем для оценки качества нейронной сети необходимо выделять 3 набора данных:

- Обучающая выборка – набор данных для обучения;

- Тестовая выборка – проверка;
- Проверочная выборка – для подбора гиперпараметров.

При проектировании искусственных нейронных сетей в основном используются парадигмы "кривых обучения". [4-7] это графики зависимости количества ошибок от количества нейронов. Анализ таких графиков позволяет определить размеры НС, необходимые для корректного функционирования.

3. Исследование готовых решений

3.1. Зависимость среднего модуля веса синапса от размера нейросети

В работе [8] вариант кривой обучения - зависимость среднего модуля веса синапса от размера нейросети.

В работе приводятся графики кривых обучения, у которых вдоль осей ординат отложено число нейронов в скрытых слоях сетей:

- средние значения ошибок обучения и обобщения (выраженные в процентах доли неправильно решенных примеров в объеме соответствующей выборки);
- среднеквадратичный вес синапса в сети - предлагаемый индикатор;
- максимум среди понейронных сумм модулей весов синапсов - индикатор

Для проведения экспериментов автор рассматривает 6 задач классификации с участием учителя. При проверки используются несколько тысяч обучающих векторов, что гарантирует отсутствие переобучения.

Как видно из экспериментальных графиков, при выборе оптимального размера сети опираться только на значение ошибки обучения недостаточно и нельзя выявить возникновение переобучения нейросети, поэтому сопоставление поведения нескольких индикаторов (как было сделано на приведенных графиках) позволяет либо более обоснованно подтвердить

выбор размера нейросети, либо увидеть возможное существование проблем (например, неадекватности модели из-за возникновения переобучения). Возможность же обойтись без проверки на тестовой выборке позволяет обучать нейросеть на всем доступном наборе примеров, без разделения его на обучающий и тестовый фрагменты, и ожидать, что с ростом числа обучающих примеров снизится и риск переобучения нейросети.

С помощью метода, предложенного в данной работе возможно достаточно надежное определение оптимального размера сети, обеспечивающего минимум ошибки обобщения. Индикатор позволяет обходиться без расчета ошибки обобщения на независимой тестовой выборке, допускает вариации путем выбора нормы и учета структурной неоднородности сети для максимизации прогностических способностей.

3.2. EagleEye

В работе [9] описывается метод EagleEye который является методом оценки подсетей для алгоритмов сжатия размера обученных нейросетей. Метод основан на адаптивной нормализации. EagleEye позволяет находить корреляции между подсетями обученной нейросети и точностью их предсказаний. С помощью EagleEye можно отобрать оптимальную структуру подсети с минимальной потерей в точности. EagleEye обходит существующие алгоритмы в задаче сжатия размера нейросети.

Визуальное представление метода, описанного в работе изображено на рисунке 6.

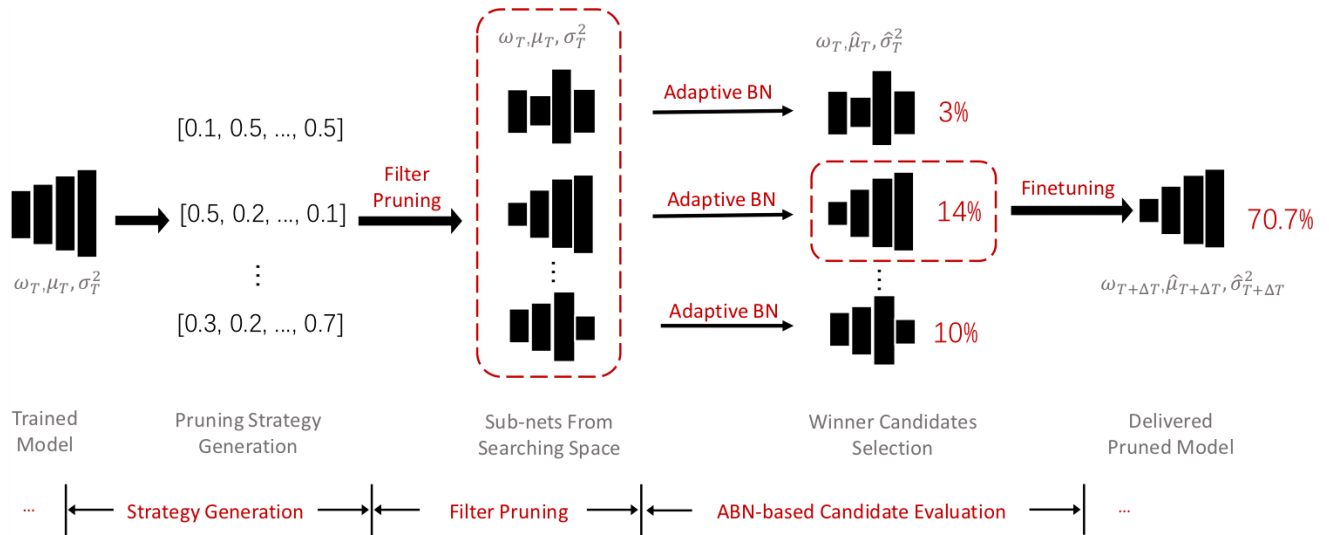


Рисунок 6 – метод EagleEye

С помощью данного подхода можно задать сеть с излишним количеством нейронов, а потом привести к более компактному решению. Такое решение является крайне эффективным так как оно дает гарантию избавления от нейронов, которые не несут в себе большой смысловой нагрузки, но при этом, они используют вычислительные способности аппарата.

Данный метод позволяет избавиться от излишних нейронов, но не дает гарантии того, что выбранная архитектура является оптимальной для данного решения.

3.3. Поиск архитектуры нейронной сети

Опишем понятия поиска архитектуры на основе информации, рассказанной в статье [1].

Поиск архитектуры нейронной сети (англ. Neural Architecture Search, NAS) — это процесс автоматизации проектирования архитектуры нейронной сети. Другими словами, это процесс поиска лучшей структуры модели машинного обучения. Система NAS получает на вход набор данных и тип задачи (классификация, регрессия и т.д.), и на выходе дает архитектуру

модели. Полученная архитектура будет работать лучше остальных архитектур для данного типа задачи при обучении на предоставленном наборе данных.

Методы для NAS классифицируются по трем категориям: пространство поиска (англ. Search Space), стратегия поиска (англ. Search Strategy) и стратегия оценки эффективности (англ. Performance Estimation Strategy). Схематичный принцип работы NAS отображен на рисунке 7.

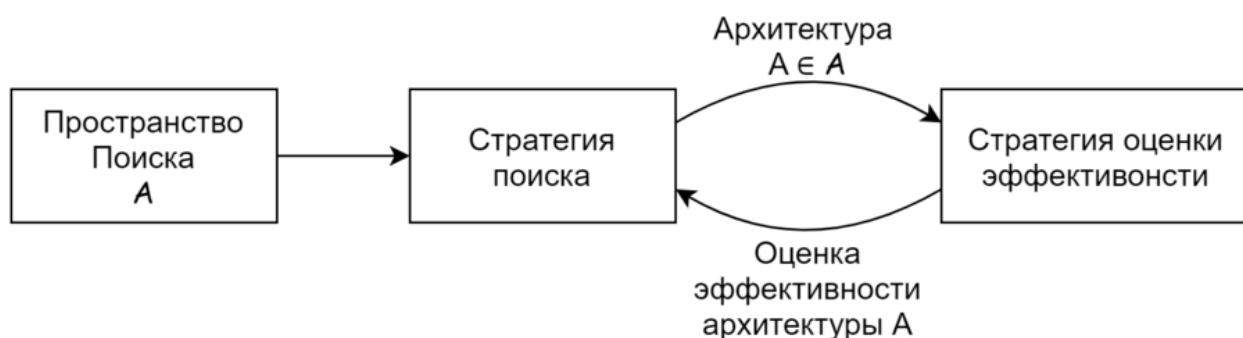


Рисунок 7 - Обобщающая иллюстрация методов NAS.

Пространство поиска определяет, какую нейронную архитектуру в принципе может обнаружить система NAS. Это может быть цепочечная архитектура, в которой выход уровня подается как вход уровня. Или это может быть сложная ветвистая архитектура с пропусками соединений.

В некоторых случаях используют спроектированный вручную каркас архитектуры (макроархитектуру), состоящий из повторяющихся ячеек (англ. motifs/blocks/cells). В таких случаях каркас является фиксированным, а задача NAS заключается в поиске архитектуры самих ячеек. Такой тип поиска известен как микро-поиск (англ. cell-search).

Схемы описанных выше архитектур изображены на рисунках 8–9.

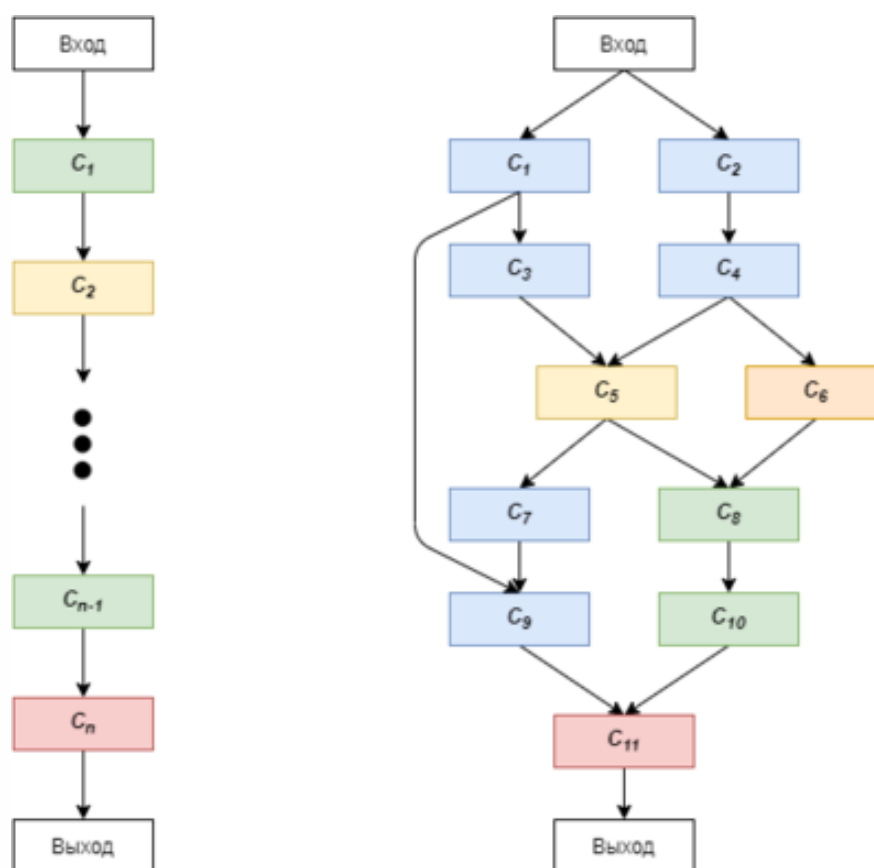


Рисунок 8 — Цепочечная и ветвистая архитектуры.

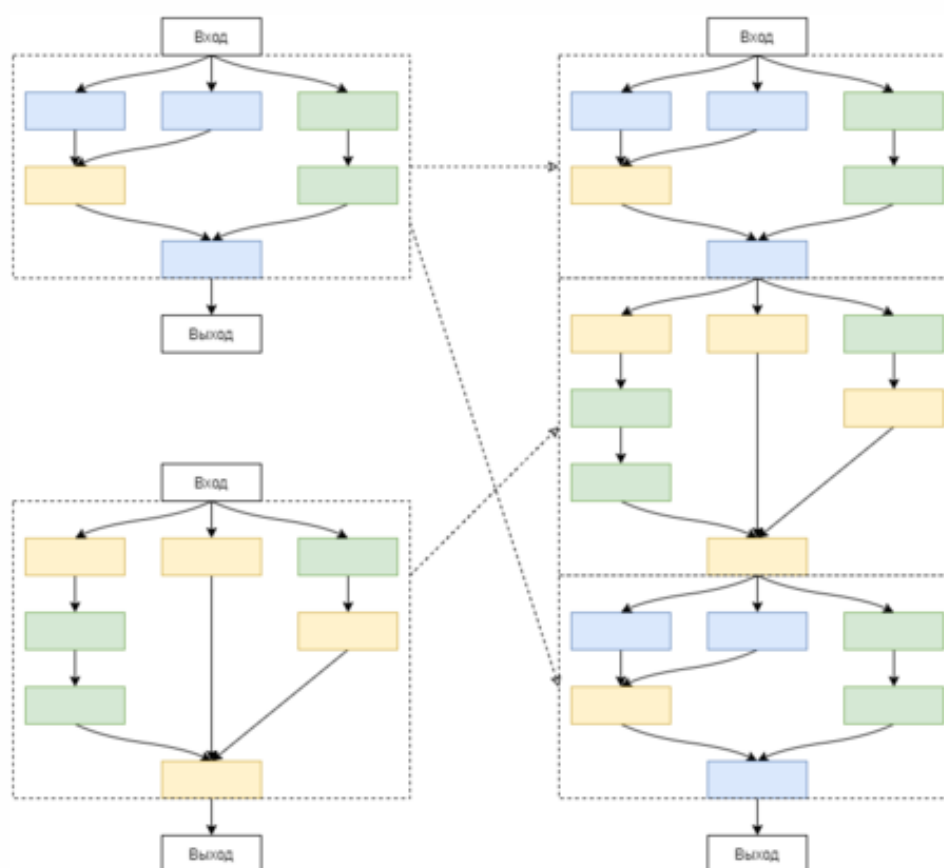


Рисунок 9 — Архитектуры ячеек.

В статье [11] приведены результаты сравнения разных методов автоматического подбора архитектуры нейронной сети. Приведем результаты сравнений на рисунке 10.

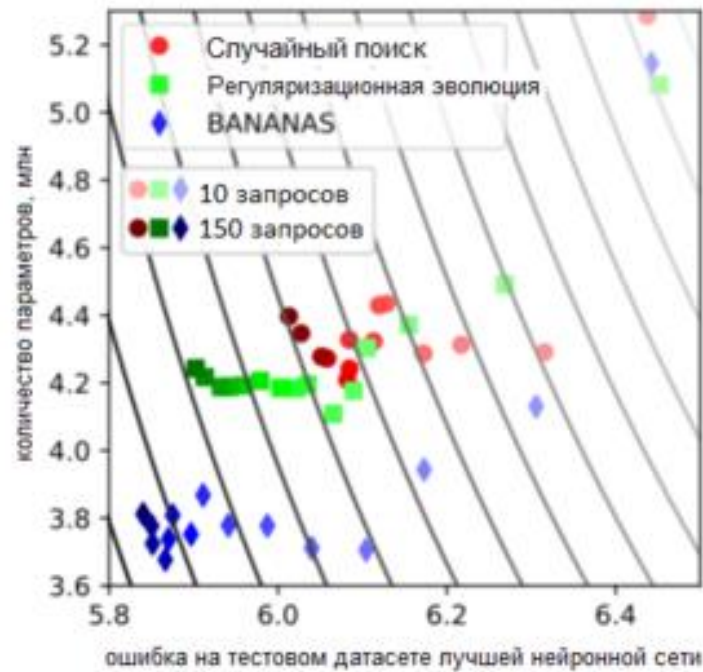


Рисунок 10 - Результат экспериментов, минимизирующих функцию потерь и количества параметров модели.

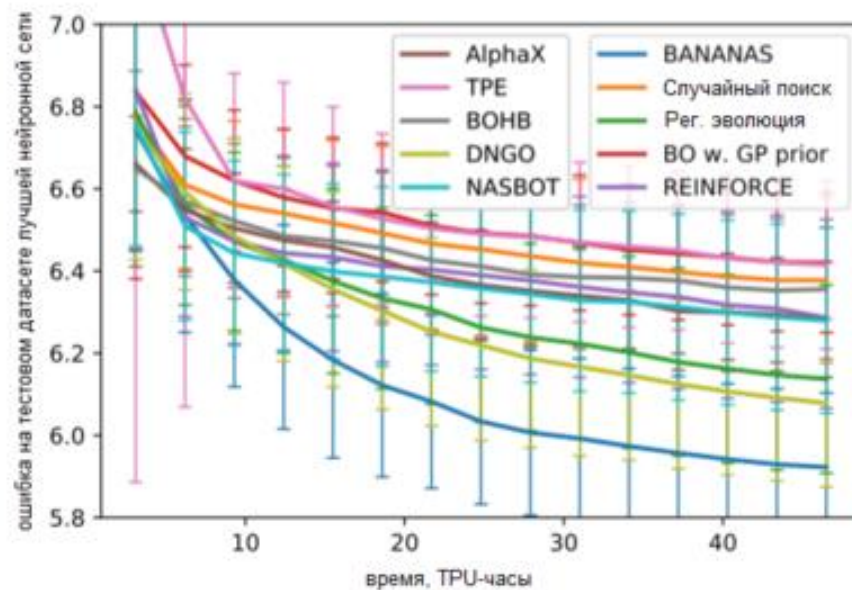


Рисунок 11 - Сравнение основных используемых в NAS алгоритмов

Вывод

В результате данной научно-исследовательской работы были изучены материалы по структуре нейронных сетей. При изучении стало понятно, что начальное количество нейронов задается с запасом, для гарантии выполнения задачи, а после с помощью алгоритма отсеиваются все недейственные части. Что касается подбора архитектуры, то так как нет четких правил используются специальные алгоритмы автоматического подбора.

Список литературы

1. Гудфеллоу Я., Бенджио И., Курвилль А. Г93 Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. Ил
2. Ксенофонов В.В./ НЕЙРОННЫЕ СЕТИ – с 28 – 29
3. Галушкин А.И. Нейронные сети: основы теории – 480 с.
4. Watanabe E., Shimizu H. Relationships between internal representation and generalization ability in multi layered neural network for binary pattern classification problem / Proc.IJCNN'1993, Nagoya, Japan, 1993. Vol.2. - pp.1736-1739.
5. Cortes C., Jackel L.D., Solla S.A., Vapnik V., Denker J.S. Learning curves: asymptotic values and rate of convergence / Advances in Neural Information Processing Systems 6 (1993). Morgan Kaufmann, 1994. - pp.327-334.
6. Cortes C., Jackel L.D., Chiang W.-P. Limits on learning machine accuracy imposed by data quality / Advances in Neural Information Processing Systems 7 (1994). MIT Press, 1995. - pp.239-246.
7. Gu B., Hu F., Liu H. Modelling classification performance for large data sets: an empirical study / Lecture Notes in Comp. Sci., 2001. Vol.2118. - pp.317-328.
8. Царегородцев В.Г, ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО РАЗМЕРА НЕЙРОСЕТИ ОБРАТНОГО РАСПРОСТРАНЕНИЯ ЧЕРЕЗ СОПОСТАВЛЕНИЕ СРЕДНИХ ЗНАЧЕНИЙ МОДУЛЕЙ ВЕСОВ СИНАПСОВ / Материалы XIV Международной конференции по нейрокибернетике, Ростов-на-Дону, 2005. Т.2. - С.60-64.
9. Bailin L, Bowen W , Jiang S , Guangrun W , Liang L, EagleEye: Fast Sub-net Evaluation for Efficient Neural Network Pruning / arXiv:2007.02491v2 [cs.CV] 5 Aug 2020 – pp. 1-16 .
10. Elsken T., Metzen J.H, Hutter F., Neural Architecture Search: A Survey/ Journal of Machine Learning Research 20 (2019) pp.1-21.

11. White C., Neiswanger W., Savani Y., BANANAS: Bayesian Optimization with Neural Architectures for Neural Architecture Search / pp. – 1 – 31