

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**ИСПОЛЬЗОВАНИЕ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНОЙ
НЕЙРОННОЙ СЕТИ В РЕШЕНИИ ЗАДАЧИ СОЗДАНИЯ ЛОГОТИПОВ**

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 273 группы
направления 02.04.03 — Математическое обеспечение и администрирование
информационных систем
факультета КНиИТ
Петрова Владимира Сергеевича

Научный руководитель

зав. каф., к. ф.-м. н., доцент

С. В. Миронов

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

Саратов 2021

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Задача генерации логотипов и пути ее решения	5
2 Описание моделей нейронных сетей, определение класса задачи	6
2.1 Нейронные сети	6
2.1.1 Основная информация	6
2.1.2 Сверточные нейронные сети	6
2.2 Генеративно-состязательные нейронные сети	7
2.2.1 Основная информация	7
2.2.2 LoGAN	8
2.2.3 Упрощенная модель	9
3 Поиск датасета, обучение моделей и анализ результатов	10
3.1 Поиск подходящего датасета	10
3.2 Обучение и результаты работы LoGAN	10
3.3 Обучение и результаты работы упрощенной модели	10
3.4 Возможные пути улучшения результатов и продолжения работы ..	11
ЗАКЛЮЧЕНИЕ	12
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	13

ВВЕДЕНИЕ

Актуальность темы

В последнее время нейронные сети находят свое применение во многих задачах. Самые классические задачи, хорошо решаемые при помощи нейронных сетей, — задачи классификации, прогнозирования, кластеризации, управления и т.д. С течением времени люди находят и применяют все более сложные модели нейронных сетей для решения более сложных задач. Одной из таких моделей является генеративно-состязательная сеть (GAN), которую описал Ян Гудфеллоу из компании Google в 2014 году [1]. Эта модель выглядит очень перспективно и уже проверена и применена на многих задачах.

Целью магистерской работы является исследование темы генеративно-состязательных нейронных сетей для реализации ПО, решающего задачу генерации логотипов. Для достижения данной цели будут решены следующие задачи:

- провести обзор основных задач, решаемых нейронными сетями;
- разобрать модель генеративно-состязательных сетей и рассмотреть типы задач, решаемых при их помощи;
- проанализировать существующие возможности реализации моделей генеративно-состязательных сетей;
- найти и разобрать нейронные сети, подходящие для решения задачи генерации логотипов;
- реализовать и обучить одну из подходящих моделей;
- проанализировать результаты и возможные дальнейшие улучшения работы.

Цель магистерской работы — исследование темы генеративно-состязательных нейронных сетей, различных реализаций моделей подобных сетей, анализ решений задач, схожих с задачей генерации логотипов и уже существующих решений поставленной задачи для реализации ПО, решающего задачу генерации логотипов.

Поставленная цель определила **следующие задачи**:

1. разобрать модель генеративно-состязательных сетей и рассмотреть типы задач, решаемых при их помощи;
2. проанализировать существующие возможности реализации моделей генеративно-состязательных сетей;

3. найти и разобрать нейронные сети, подходящие для решения задачи генерации логотипов или смежных задач;
4. реализовать и обучить одну из подходящих моделей;
5. проанализировать результаты обучения и возможные дальнейшие улучшения работы.

Методологические основы использования генеративно-сопоставительных нейронных сетей в решении задачи создания логотипов и схожих задач представлены в работах Goodfellow, Luc, Brownlee, Sage, Agustsson, Timofte, и Van Gool.

Практическая значимость магистерской работы

Данная работа является значимой, так как задача генерации логотипов при помощи генеративно-сопоставительных сетей еще мало изучена и в этой сфере можно придумать множество решений, генерирующих хорошие результаты. Логотипы — важная часть любой организации или веб-сайта. Они зачастую являются вторым по запоминаемости объектом (после названия), поэтому важно уметь подобрать красивый и запоминающийся логотип. Автоматизация решения данной задачи может сэкономить много времени дизайнерам, что является еще одной причиной значимости данной работы.

Структура и объем работы

Магистерская работа состоит из введения, трех разделов, заключения, списка использованных источников и 12 приложений. Общий объем работы — 112 страниц, из них 57 страниц — основное содержание, включая 37 рисунков, список использованных источников информации — 36 наименований.

1 Задача генерации логотипов и пути ее решения

Этот раздел посвящен описанию термина «логотип», обозначению его важности, а также формулировке задачи генерации логотипов и обозначению решения этой задачи в магистерской работе.

Термин «логотип» обычно применяется к любым графическим и/или шрифтовым знакам, символам и эмблемам, которые могут принадлежать фирмам, организациям, проектам и событиям, отдельным людям, товарам и услугам. Важность этого, казалось бы, узкоспециального дизайнерского жанра именно для веб-дизайна объясняется просто: большинство страниц и сайтов в современном Интернете совсем не нуждаются в обильной, «агрессивной» рекламе, а многие некоммерческие сайты просто не могут позволить себе иметь такого рода оформление. Поэтому не что иное как логотип очень часто служит основой, главным (а на страницах, оформленных в академическом стиле, нередко и единственным) графическим элементом веб-страницы [2].

Задача, решаемая в настоящей работе — задача создания логотипов. Эта задача является довольно актуальной на данный момент времени, так как при основании практически любой фирмы требуется создание подходящего для нее, уникального и привлекательного логотипа. Обычно этой работой занимаются дизайнеры, но для получения хорошего логотипа обычно требуется потратить довольно много времени и денег. Если решать эту задачу автоматизированно (при условии того, что результат будет достаточно подходящим), можно будет избавиться по меньшей мере от траты большого количества времени.

В этой работе данная задача будет решаться при помощи генеративно-состязательной нейронной сети. Далее будет подробно описано, что это за модель и почему будет использоваться именно она. Пока же можно сказать, что генеративно-состязательные сети очень хорошо умеют «создавать» что-либо или же приближать к реалистичным какие-либо наброски, поданные на вход.

Таким образом, был обозначен термин «логотип», его важность, перечислены основные решения задачи генерации логотипов и описан способ решения данной задачи в магистерской работе.

2 Описание моделей нейронных сетей, определение класса задачи

Этот раздел посвящен обзору теоретических материалов, необходимых для создания модели генеративно-состязательной нейронной сети, подходящей для решения задачи генерации логотипов, а также обзору существующих решений данной задачи с использованием генеративно-состязательных нейронных сетей.

2.1 Нейронные сети

2.1.1 Основная информация

Искусственная нейронная сеть (ИНС, в дальнейшем под нейронной сетью будет подразумеваться именно ИНС) — математическая модель, а также ее программная или аппаратная реализация, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма. Это понятие возникло при изучении процессов, протекающих в мозге, и при попытке описать и компьютеризировать эти процессы. Первой такой попыткой были простейшие нейронные сети У. Маккалока и У. Питтса [3], именно они и заложили основы искусственного интеллекта и одними из первых увидели представление мозга как компьютера, что в дальнейшем стимулировало развитие кибернетики, теоретической нейрофизиологии, а также компьютерных наук [4].

2.1.2 Сверточные нейронные сети

Сверточная нейронная сеть (convolutional neural network, CNN) — специальная архитектура искусственных нейронных сетей, предложенная Яном Лекуном в 1988 году и нацеленная на эффективное распознавание образов. Использует некоторые особенности зрительной коры, в которой были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определенного набора простых клеток. Основная идея таких нейронных сетей заключается в чередовании сверточных слоев (convolution layers) и субдискретизирующих слоев (subsampling layers), также называемых слоями подвыборки (pooling layers). Структура сети — однонаправленная, многослойная. Обучается стандартными методами (такими, как метод обратного распространения ошибки, используется чаще всего). Название данная архитектура получила из-за наличия операции свертки, суть которой в том, что каждый фрагмент

изображения умножается на матрицу (ядро) свертки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.

Основная идея работы сверточной нейронной сети заключается в переходе от конкретных особенностей изображения к более абстрактным деталям, и далее к еще более абстрактным деталям вплоть до выделения понятий высокого уровня. При этом сеть самонастраивается и вырабатывает сама необходимую иерархию абстрактных признаков (последовательности карт признаков), фильтруя маловажные детали и выделяя существенные.

2.1.2.1 Структура сверточной нейронной сети

В сверточной нейронной сети выходы промежуточных слоев образуют матрицу (изображение) или набор матриц (несколько слоев изображения). Так, например, на вход сверточной нейронной сети можно подавать три слоя изображения (R-, G-, B-каналы изображения). Основными видами слоев в сверточной нейронной сети являются сверточные слои (convolutional layer), пулинговые слои (pooling layer) и полносвязные слои (fully-connected layer).

Сверточный слой нейронной сети представляет из себя применение операции свертки к выходам с предыдущего слоя, где веса ядра свертки являются обучаемыми параметрами. Еще один обучаемый вес используется в качестве константного сдвига (англ. bias).

Пулинговый слой призван снижать размерность изображения. Исходное изображение делится на блоки размером $w \times h$ и для каждого блока вычисляется некоторая функция. Чаще всего используется функция максимума (max pooling) или (взвешенного) среднего ((weighted) average pooling). Обучаемых параметров у этого слоя нет. Основные цели пулингового слоя:

- уменьшение изображения, чтобы последующие свертки оперировали над большей областью исходного изображения;
- увеличение инвариантности выхода сети по отношению к малому переносу входа;
- ускорение вычислений.

2.2 Генеративно-сопоставительные нейронные сети

2.2.1 Основная информация

Генеративно-сопоставительная сеть (англ. Generative adversarial network, сокращённо GAN) — алгоритм машинного обучения без учителя, построенный

на комбинации из двух нейронных сетей, одна из которых (сеть G) генерирует образцы, а другая (сеть D) старается отличить правильные («подлинные») образцы от неправильных. Так как сети G и D имеют противоположные цели — создать образцы и отбраковать образцы — между ними возникает антагонистическая игра. Генеративно-сопоставительную сеть описал Ян Гудфеллоу [1] из компании Google в 2014 году.

Работает эта сеть следующим образом: используя набор переменных латентного пространства, генеративная сеть (сеть G) пытается «слепить» новый образец, смешав несколько исходных образцов. Дискриминативная сеть (сеть D) обучается различать подлинные и поддельные образцы, а результаты различения подаются на вход генеративной сети так, чтобы она смогла подобрать лучший набор латентных параметров [5], и дискриминативная сеть уже не смогла бы отличить подлинные образцы от поддельных. Таким образом целью сети G является повысить процент ошибок сети D, а целью сети D является наоборот улучшение точности распознавания.

2.2.2 LoGAN

В процессе поиска подходящей для данной задачи модели сначала было решено взять модель LoGAN [6] — одну из первых в мире нейросетей, решающих схожие задачи.

В работе «Logo synthesis and manipulation with clustered generative adversarial network» [7] A. Sage, E. Agustsson, R. Timofte, и L. Van Gool исследовали возможное решение задачи генерации логотипов. Они предложили кластерный подход, работающий с мультимодальными данными, в особенности, с логотипами. Каждому логотипу назначается метка, которая определяет, к какому кластеру принадлежит логотип, и генеративно-сопоставительная сеть тренируется, принимая во внимание эти самые метки.

Для описания модели, использующейся в настоящей работе, необходимо рассмотреть две модификации генеративно-сопоставительной нейронной сети, называющиеся Conditional GAN и Auxiliary Classifier GAN.

Conditional GAN: в conditional GAN (CGAN) [8] дискриминатор и генератор имеют также условие c , которое может быть либо меткой класса, либо другими данными подобного рода. Входные данные и c комбинируются в единое представление информации и подаются на вход как дополнительный слой в обеих сетях.

Auxiliary Classifier GAN (ACGAN): в отличие от CGAN, внутри ACGAN [9] латентное пространство z содержит внутри себя условия меток класса (то есть метки находятся не в изображениях, а в самом пространстве). Дискриминатор в этом случае должен уметь как отличать настоящие изображения от реальных, так и определять класс объекта, вне зависимости от того, настоящий он или нет.

В настоящей работе использовалась немного модифицированная версия ACGAN: предложенной архитектурой для LoGAN является Auxiliary Classifier Wasserstein Generative Adversarial Neural Network с градиентным штрафом (AC-WGAN-GP), которая является модификацией Auxiliary Classifier GAN.

2.2.3 Упрощенная модель

После исследования модели AC-WGAN-GP оказалось, что она довольно избыточная в рамках исследовательской работы. Она содержит в себе огромное количество лишнего кода, загружает весь датасет в оперативную память, что не позволяет обучать ее на компьютерах с небольшим количеством ОЗУ. Также содержит метки классов, которые сильно усложняют саму сеть и увеличивают сложность обучения сети. Из-за всех этих недостатков было принято решение использовать буквально «чистую» генеративно-сопоставительную сеть, состоящую из довольно простых моделей классификатора и дискриминатора для достижения основной цели — генерации логотипов.

Модель генератора работает по очень простому принципу — она получает на вход вектор случайного шума заданной длины, а затем начинает его интерполировать по слоям нейронной сети до тех пор, пока не получится изображение необходимого размера. Функцией потерь для генератора была выбрана перекрестная энтропия (кросс-энтропия).

Модель дискриминатора является по сути обычным классификатором на основе CNN [10].

Таким образом, была приведена основная теория, необходимая для создания модели, подходящей для решения задачи генерации логотипов, а также было описано уже существующее решение данной задачи при помощи генеративно-сопоставительной нейронной сети.

3 Поиск датасета, обучение моделей и анализ результатов

Данный раздел посвящен поиску датасета, необходимого для обучения модели, обучению самой модели и анализу полученных результатов.

3.1 Поиск подходящего датасета

В работе [7] описывается создание так называемого LLD — Large Logo Dataset, который был создан специально для обучения нейронных сетей, которые должны генерировать логотипы.

В основном все научные исследования были сфокусированы на опознании существующих логотипов (задача классификации) [11–16] и, конечно же, для обучения этих нейронных сетей уже были собраны некоторые датасеты. Из-за небольшой разнообразности содержащихся там данных эти датасеты не очень хорошо подходят для автоматизированной генерации логотипов. В данном же датасете (LLD) было собрано огромное количество разнообразных иконок и логотипов.

3.2 Обучение и результаты работы LoGAN

Нейронная сеть обучалась на ноутбуке DELL G3 3579 с процессором Intel Core I5-8300H 2.30GHz, 8GB RAM и видеокартой NVIDIA GTX 1050.

К сожалению, из-за внутренних недостатков взятой за основу AC-WGAN-GP, обучение прошло только 300 эпох на 100 тысячах изображений. В качестве данных для обучения были выбраны иконки размера 32×32 , но существует возможность обучения сети на иконках большего размера (и большем количестве), либо же использование другой генеративно-сопоставительной сети для улучшения качества получившихся логотипов. Поиск или создание лучшего датасета — довольно трудоемкая задача, потому что LLD на данный момент считается лучшим датасетом, содержащим различные логотипы. Поэтому впоследствии было принято решение обучать другую нейронную сеть.

Данная сеть не была сильно модифицирована относительно опубликованной в работе [6] реализации и имеет некоторые недостатки (например, загрузка всего датасета сразу и долгий его препроцессинг).

3.3 Обучение и результаты работы упрощенной модели

Как было упомянуто выше, упрощенная модель нейронной сети, решающей задачу генерации логотипов, стала классическая реализация, состоящая

из простого классификатора, основанного на сверточной нейронной сети, и генератора, который работает с точностью наоборот (вместо сжатия изображения для выделения признаков вектор случайного шума, являющийся набором признаков, интерполируется в изображение).

Нейронная сеть обучалась на персональном компьютере с процессором Intel Core I5-9600K 3.7GHz, 32GB RAM и видеокартой NVIDIA RTX 3070.

Для обучения нейронной сети на графическом процессоре использовалась NVIDIA CUDA 11.0 с библиотекой cuDNN версии 8.0.5. Из-за современности графического процессора необходимо было взять компилятор ptxas из библиотеки NVIDIA CUDA 11.1, так как его версия в 11.0 не могла корректно работать с видеокартой.

Данная сеть обучалась на 100 тысячах изображений, пройдя 5050 эпох. В качестве данных для обучения были выбраны иконки размера 32×32 . Тем не менее, существует возможность модифицировать данную сеть так, чтобы она смогла обучаться на иконках большего размера для получения результатов в более высоком разрешении. Обучение заняло порядка 84 часов, результатами обучения являются размытые фигуры различной формы и цветов. Позже была увеличена резкость результатов посредством применения свертки к полученным изображениям.

3.4 Возможные пути улучшения результатов и продолжения работы

Данная работа при небольших ресурсах имеет достаточно неплохие результаты, которые тем не менее можно сильно улучшить и сделать их более настраиваемыми. Можно предложить следующий список улучшений, которые можно реализовать при продолжении работы:

- расширение размера датасета;
- упрощение или усложнение модели;
- добавление кастомизации (настраиваемости) в нейронную сеть;
- улучшение качества результирующих изображений.

Таким образом, в данном разделе было приведено описание датасета, используемого для обучения модели нейронной сети, а также приведены данные обучения всех использованных моделей, краткое описание результатов работы и возможные пути улучшения решения.

ЗАКЛЮЧЕНИЕ

В ходе выполнения данной работы была сформулирована задача, которая решалась в ходе выполнения магистерской работы, также была обоснована актуальность данной работы. Был приведен обширный обзор теории, необходимой для понимания общего решения задачи, поиска, реализации и обучения подходящих моделей нейронной сети. Были проанализированы смежные задачи и существующие решения поставленной в магистерской работе задачи. Также был найден подходящий для обучения датасет (впоследствии приспособленный под обучение двух различных генеративно-состязательных нейронных сетей). Были реализованы две различные модели генеративно-состязательной нейронной сети, которые решают поставленную задачу. Было проведено сравнение эффективности обучения и получившихся результатов. Результаты работы были проанализированы и улучшены, также были предложены пути продолжения и улучшения результатов данной работы.

Подводя итог, в данной работе было показано, как искусственный интеллект на примере генеративно-состязательных сетей может помочь дизайнерам в процессе создания логотипов. Предложенная модель может создавать различные случайные логотипы в небольшом разрешении. Вследствие этого они могут использоваться только как наброски для настоящих работ, что в любом случае может сохранить некоторое время работы дизайнеров.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Goodfellow, I. J.* Generative adversarial nets // Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2. — NIPS'14. — Cambridge, MA, USA: MIT Press, 2014. — Pp. 2672–2680.
- 2 *Дмитрий, К.* Веб-дизайн / К. Дмитрий. — Россия, Санкт-Петербург: СИМВОЛ-Плюс, 2006.
- 3 *McCulloch, W. S.* A logical calculus of the ideas immanent in nervous activity / W. S. McCulloch, W. Pitts // *The bulletin of mathematical biophysics*. — Dec 1943. — Vol. 5, no. 4. — Pp. 115–133. <https://doi.org/10.1007/BF02478259>.
- 4 *Smalheiser, N. R.* Walter pitts / N. R. Smalheiser // *Perspectives in Biology and Medicine*. — 2000. — Vol. 43, no. 2. — Pp. 217–226.
- 5 *Luc, P.* Semantic segmentation using adversarial networks / P. Luc, C. Couprie, S. Chintala, J. Verbeek // *arXiv preprint arXiv:1611.08408*. — 2016.
- 6 *Mino, A.* Logan: Generating logos with a generative adversarial neural network conditioned on color / A. Mino, G. Spanakis // *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. — Dec 2018. <http://dx.doi.org/10.1109/ICMLA.2018.00157>.
- 7 *Sage, A.* Logo synthesis and manipulation with clustered generative adversarial networks / A. Sage, R. Timofte, E. Agustsson, L. V. Gool // *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. — Jun 2018. <http://dx.doi.org/10.1109/CVPR.2018.00616>.
- 8 *Mirza, M.* Conditional generative adversarial nets. arxiv / M. Mirza, S. Osindero // *Learning*. — 2014. — Vol. 2, no. 4.2.
- 9 *Odena, A.* Conditional image synthesis with auxiliary classifier GANs // Proceedings of the 34th International Conference on Machine Learning / Ed. by D. Precup, Y. W. Teh. — Vol. 70 of *Proceedings of Machine Learning Research*. — International Convention Centre, Sydney, Australia: PMLR, 06–11 Aug 2017. — Pp. 2642–2651. <http://proceedings.mlr.press/v70/odena17a.html>.

- 10 *O'Shea, K.* An introduction to convolutional neural networks / K. O'Shea, R. Nash // *CoRR*. — 2015. — Vol. abs/1511.08458. <http://arxiv.org/abs/1511.08458>.
- 11 *Hoi, S. C. H.* Logo-net: Large-scale deep logo detection and brand recognition with deep region-based convolutional networks / S. C. H. Hoi, X. Wu, H. Liu, Y. Wu, H. Wang, H. Xue, Q. Wu // *CoRR*. — 2015. — Vol. abs/1511.02462. <http://arxiv.org/abs/1511.02462>.
- 12 *Joly, A.* Logo retrieval with a contrario visual query expansion // Proceedings of the 17th ACM International Conference on Multimedia. — MM-09. — New York, NY, USA: Association for Computing Machinery, 2009. — Pp. 581–584. <https://doi.org/10.1145/1631272.1631361>.
- 13 *Romberg, S.* Scalable logo recognition in real-world images // Proceedings of the 1st ACM International Conference on Multimedia Retrieval. — ICMR-11. — New York, NY, USA: Association for Computing Machinery, 2011. — 8 pp. <https://doi.org/10.1145/1991996.1992021>.
- 14 *Sahbi, H.* Context-dependent logo matching and recognition / H. Sahbi, L. Ballan, G. Serra, A. Del Bimbo // *IEEE Transactions on Image Processing*. — 2013. — Vol. 22, no. 3. — Pp. 1018–1031.
- 15 *Su, H.* Weblogo-2m: Scalable logo detection by deep learning from the web // 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). — 2017. — Pp. 270–279.
- 16 *Zhu, G.* Automatic document logo detection // Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). — Vol. 2. — 2007. — Pp. 864–868.