

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

Отчет о командном исследовательском проекте на тему: Исследование влияния *inductive bias* на обучение нейронных сетей

Выполнили:

студент группы БПМИ223 Леонтьев Константин Валерьевич
студент группы БПМИ223 Тань Сипэн

Принял руководитель проекта:

Болдырев Алексей Сергеевич
Доцент Департамента больших данных и информационного поиска ФКН

Москва 2025

Содержание

Аннотация	4
1 Введение	5
1.1 Inductive bias и обучение	5
1.2 Attention как вид Inductive Bias	6
1.3 Примеры	6
1.4 Задачи и цели проекта	6
2 Постановка задачи	7
3 Данные и датасеты	8
3.1 Food-101	8
3.2 Mini-ImageNet	8
3.3 DomainNet	9
3.4 Dogs vs Cats	9
4 Модели	10
4.1 Модель для Food-101	10
4.1.1 Без Inductive Bias	10
4.1.2 С Inductive Bias	11
4.1.3 Визуализация Attention	12
4.2 Модель для Mini-ImageNet	13
4.2.1 Без Inductive Bias	13
4.2.2 С Inductive Bias	13
4.2.3 Визуализация Attention	14
4.3 Модель для DomainNet и Dogs vs. Cats	14
5 Обучение	15
6 Результаты	17
6.1 Food-101	17
6.1.1 До внедрения Inductive Bias	17
6.1.2 После внедрения Inductive Bias	17
6.1.3 Сравнение с другими бенчмарками	18
6.2 Mini-ImageNet	19

6.2.1	До внедрения Inductive Bias	19
6.2.2	После внедрения Inductive Bias	20
6.2.3	Сравнение с другими бенчмарками	21
6.3	DomainNet	21
6.4	Dogs vs Cats	21
7	Итоги	23
8	Обзор литературы	24
	Список литературы	25

Аннотация

В данной работе исследуется влияние априорных допущений (inductive bias) на обучение нейронных сетей. Когда мы обучаем модель, мы делаем какие-то априорные предположения об устройстве наших данных, об архитектуре, которая подошла бы для этой задачи лучшим образом. Например, когда мы обучаем линейную модель, мы предполагаем, что есть какая-то линейная зависимость между признаками и целевой переменной. Однако, чаще всего мы не знаем этого. Наша задача заключается в том, чтобы понять как именно inductive bias влияет на процесс обучения модели и её обобщающую способность.

Ключевые слова

Inductive bias, нейронные сети, глубинное обучение, машинное обучение, обобщающая способность модели

1 Введение

Inductive bias (IB) представляет собой априорные допущения о природе данных, вложенные в модель машинного обучения. IB позволяет указать модели на предпочтительный способ генерализации в условиях обучающей выборки конечного размера. Иными словами — это предположения или ограничения, которые модель машинного обучения или алгоритм использует, чтобы обобщать знания на новых данных. Например, человек может с помощью небольших данных быстро найти закономерность (обобщение на основе опыта, эвристики) и при встрече с новыми задачами может получить более хорошие результаты чем модель. Как и у людей, у машин есть индуктивное смещение, которое помогает им «учиться» из ограниченного количества данных.

1.1 Inductive bias и обучение

Однако не стоит путать *inductive bias* и процесс обучения, так как это принципиально разные вещи. Индуктивное смещение включает в себя подбор подходящей архитектуры, выбор регуляризации и функции потери, генерацию новых признаков, отбор старых признаков, подбор гиперпараметров. В то время как обучение заключается в подборе подходящих параметров, минимизации приходящей функции потерь и то, каким способом мы это делаем. Без индуктивного смещения процесс обучения был бы невозможен. То есть IB — задание правил игры, а обучение — это следование этим правилам.

Таким образом, решение любой задачи по глубинному/машинному обучению можно свести к следующим шагам:

- *Формулировка задачи* — На этом этапе определяется цель обучения модели и тип данных, с которыми мы будем работать.
- *Выбор модели* — Самый важный этап, так как тип модели тесно связано с задачей.
- *Выбор гиперпараметров* — Например подбор регуляризации, размера модели, глубины дерева решений.
- *Обучение модели* - Во время обучения IB проявляется в способе, которым модель оптимизирует свои параметры: оптимизации, функции потерь.
- *Обобщение на новых данных* — Если модель обучена на изображениях объектов, но видит новые изображения, например анимешные изображения тех же объектов, и нужно выяснить достаточное ли IB было сделано.

1.2 Attention как вид Inductive Bias

Сам по себе attention не является сильным inductive bias, особенно в глубоких нейронных сетях. Чаще всего он деёт весовой результат например с такими механизмами как positional encoding. Однако в моделях послабее он может придать какое-то определённое улучшение и увеличить качество. Attention — один из основных механизмов inductive bias, однако он дает лучшие показатели именно в совокупе с другими частями архитектуры.

1.3 Примеры

Существует множество примеров, когда индуктивный сдвиг играет огромную роль в обучении. Например, от него может зависеть в какой момент (на какой эпохе) нашего алгоритма loss начнет резко снижаться (Grokking): на одной архитектуре он может произойти на 50-й эпохе, а на другой на 5000-й, и разница тут весомая.

Ещё один пример, показывающий важность inductive bias, это трансферное обучение. Ведь здесь от выбранной архитектуры зависит насколько велика генерализующая способность модели.

Inductive bias — очень обширная, фундаментальная и актуальная задача машинного обучения, которая требует дальнейшего исследования.

1.4 Задачи и цели проекта

В задачи нашего проекта входит:

- 1 Изучение методов работы Inductive Bias
- 2 Выбор задачи и построение модели на выбранном датасете
- 3 Сравнение архитектур между собой
- 4 Сравнение результатов с результатами из опубликованных исследований и подведение итогов

2 Постановка задачи

Для того, чтобы наглядно показать насколько сильно Inductive Bias влияет на процесс обучения, будем рассматривать задачу многоклассовой классификации на примере небольших нейронных сетей. Такой подход поможет более наглядно отобразить разницу результатов обучения с и без внедрения Inductive Bias.

Нами было принято решение использовать обычные свёрточные нейронные сети (CNN). Простота этих нейронных сетей несёт внутри себя огромный потенциал, который можно ещё больше раскрыть при помощи внедрения Inductive Bias.

Также, хотелось показать, насколько велика обобщающая способность нашей модели с внедрением Inductive Bias и без него. И сделать некие выводы о природе наших данных.

3 Данные и датасеты

Для наших экспериментов были выбраны несколько датасетов: Food-101, Mini-ImageNet, DomainNet и Dogs vs. Cats. Все они подходят для решения задачи классификации.

3.1 Food-101

Этот датасет [7] состоит из набора классов различной еды. Всего он содержит 101 класс, на каждый из которых приходится по 750 изображений для обучения и 250 изображений для теста. Итого, общее количество изображений составляет 101 000. В примечаниях к этому датасету было написано, что в обучающей части содержится небольшой шум в данных.



Рис. 3.1: Food-101

3.2 Mini-ImageNet

Данный датасет [10] состоит из 100 классов. Общее количество изображений — 60 000, из которых 50 000 отложены под обучающую выборку, а остальные 10 000 — под тестовую. Сам датасет представляет собой уменьшенную версию ImageNet, с урезанным количеством классов и разрешением изображений 64×64 .



Рис. 3.2: Mini-ImageNet

3.3 DomainNet

DomainNet [6] — это набор данных в шести различных доменах. Все домены включают 345 категорий (классов) объектов, таких как браслет, самолет, птица, виолончели и т.п. Домены включают в себя клипарты; реальные фотографии и изображения; эскизы; инфографику; рисунки художественных изображений объектов в виде картин и рисунки от руки.

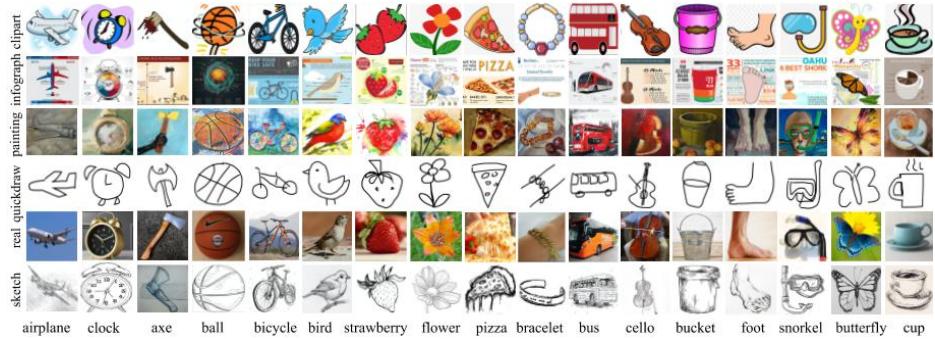


Рис. 3.3: DomainNet

3.4 Dogs vs Cats

Dogs vs. Cats [5] — это набор данных, состоящий из изображений кошек и собак в двух стилях: реальные фотографии и рисунки от руки. Задача — бинарная классификация. Модель обучается на реальных фотографиях, а классификация происходит на рисунках от руки, что позволяет оценить обобщающую способность модели на изображениях с другим стилем.

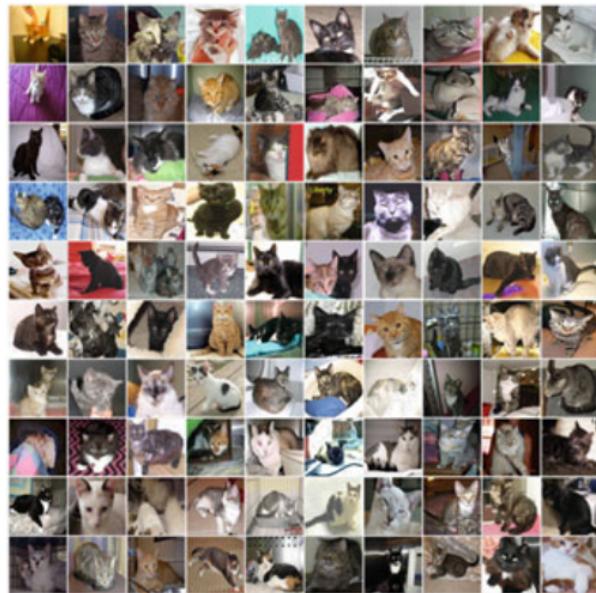


Рис. 3.4: Dogs vs. Cats

4 Модели

В основе наших экспериментов лежит внедрение Spatial Attention внутрь архитектуры модели (это и есть наш Inductive Bias). В таблице можно более детально изучить слои данного вида attention.

Номер слоя	Тип слоя	Параметры
1	Input	
2	Mean Pool	
3	Max Pool	
4	Concatenate	
5	Conv2d	n_channels=2, out_channels=1
6	Sigmoid	
7	Multiply	

Таблица 4.1: Spatial Attention

Этот вид Attention предназначен для выделения определённых участков изображения. Таким образом в дальнейшем мы будем выделять отсюда признаки и вставлять внутрь полно связного слоя в нашей модели.

4.1 Модель для Food-101

4.1.1 Без Inductive Bias

Была выбрана обычная нейронная сеть, состоящая из нескольких свёрточных слоёв. Модель изображена в таблице ниже.

№ слоя/блока	Тип слоя	Параметры
1	Input	
2	Sequential Conv2d BatchNorm2d ReLU Conv2d BatchNorm2d	in_channels=3, out_channels=32, kernel_size=3 num_features=32 in_channels=32, out_channels=32, kernel_size=3 num_features=32
3	Conv2d	n_channels=3, out_channels=32, kernel_size=3
4	Addition	сумма слоя 2 и 3
5	Sequential ReLU AvgPool2d	kernel_size=8
6	Flatten	
7	Linear	in_features=8192, out_features=128
8	Linear	in_features=128, out_features=101

Таблица 4.2: CNN для Food-101

В данной модели содержится около 1М параметров, что является небольшим относительно, например, других моделей, которые использовались другими исследователями для решения задачи классификации на данном датасете.

4.1.2 C Inductive Bias

№ слоя/блока	Тип слоя	Параметры
1	Input	
2	Sequential Conv2d BatchNorm2d ReLU Conv2d BatchNorm2d	in_channels=3, out_channels=32, kernel_size=3 num_features=32 in_channels=32, out_channels=32, kernel_size=3 num_features=32
3	Conv2d	n_channels=3, out_channels=32, kernel_size=3
4	Addition	сумма слоя 2 и 3
5	Sequential ReLU AvgPool2d	kernel_size=8
6	Flatten	
7	Linear	in_features=8192, out_features=128
8	ReLU	

Таблица 4.3: Часть 1 с Inductive Bias для Food-101

№ слоя/блока	Тип слоя	Параметры
1	Input	
2	Sequential Conv2d BatchNorm2d ReLU Conv2d BatchNorm2d	in_channels=3, out_channels=32, kernel_size=3 num_features=32 in_channels=32, out_channels=32, kernel_size=3 num_features=32
3	Conv2d	n_channels=3, out_channels=32, kernel_size=3
4	Addition	сумма слоя 2 и 3
5	SpatialAttention	kernel_size=7
6	Sequential Conv2d BatchNorm2d ReLU SpatialAttention AvgPool2d Flatten Linear ReLU	in_channels=32, out_channels=64, kernel_size=3 num_features=64 kernel_size=7 kernel_size=4 in_features=1024, out_features=512

Таблица 4.4: Часть 2 с Inductive Bias для Food-101

№ слоя/блока	Тип слоя	Параметры
1	Concatenate	Конкatenируем выходы из части 1 и 2
2	Linear	in_features=1024 + 512, out_features=101

Таблица 4.5: Часть 3 с Inductive Bias для Food-101

Таким образом мы добавляем дополнительные признаки перед полно связанным слоем. Это действие и является у нас Inductive Bias.

В этой модели уже больше параметров (9М), однако это всё равно достаточно небольшие размеры по сравнению с другими моделями, обученными на этом датасете.

4.1.3 Визуализация Attention

Мы визуализировали наш Attention слой. Он подсвечивает самые важные регионы картинки, которые хорошо могут сработать в качестве дополнительных признаков.

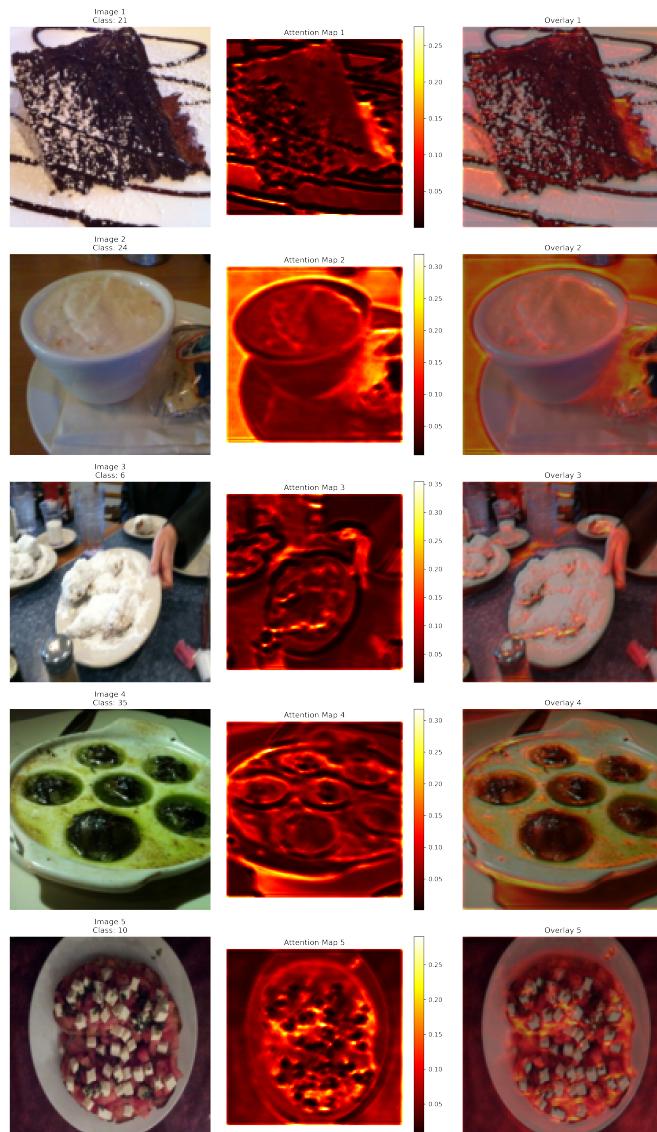


Рис. 4.1: Attention для Food-101

4.2 Модель для Mini-ImageNet

4.2.1 Без Inductive Bias

Здесь была выбрана почти такая же модель как и для Food-101, с небольшими изменениями параметров.

№ слоя/блока	Тип слоя	Параметры
1	Input	
2	Sequential Conv2d BatchNorm2d ReLU Conv2d BatchNorm2d	in_channels=3, out_channels=64, kernel_size=3 num_features=64 in_channels=64, out_channels=64, kernel_size=3 num_features=64
3	Conv2d	n_channels=3, out_channels=64, kernel_size=3
4	Addition	сумма слоя 2 и 3
5	Sequential ReLU AvgPool2d	kernel_size=8
6	Flatten	
7	Linear	in_features=4096, out_features=100

Таблица 4.6: CNN для Mini-ImageNet, количество параметров: 449188

4.2.2 С Inductive Bias

№ слоя/блока	Тип слоя	Параметры
1	Input	
2	Sequential Conv2d BatchNorm2d ReLU Conv2d BatchNorm2d ReLU AvgPool2d	in_channels=3, out_channels=64, kernel_size=3 num_features=64 in_channels=64, out_channels=64, kernel_size=3 num_features=64 kernel_size=2
3	SpatialAttention	kernel_size=7
4	Sequential Conv2d BatchNorm2d ReLU Conv2d BatchNorm2d ReLU AvgPool2d	in_channels=64, out_channels=128, kernel_size=3 num_features=128 in_channels=128, out_channels=128, kernel_size=3 num_features=128 kernel_size=2

5	SpatialAttention	kernel_size=7
6	Sequential Conv2d BatchNorm2d ReLU Conv2d BatchNorm2d ReLU AvgPool2d	in_channels=128, out_channels=256, kernel_size=3 num_features=256 in_channels=256, out_channels=256, kernel_size=3 num_features=256 kernel_size=2
7	SpatialAttention	kernel_size=7
8	AdaptiveAvgPool2d	output_size=(1, 1)
9	Flatten	
10	Linear	in_features=256, out_features=100

Таблица 4.7: CNN с Inductive Bias для Mini-ImageNet, количество параметров: 1173197

Inductive Bias добавлен в модель через добавление специальных attention-модулей, эти Spatial Attention определяют важные расположение картинки.

4.2.3 Визуализация Attention

Мы визуализировали наш Attention слой для Mini-ImageNet. Он подсвечивает ряд важных частей картинки: модель сосредотачивается на более информативных частях изображения, например, для изображения птицы светло выделены головы и клюва, что и соответствует признакам птиц с точки зрения классификации.

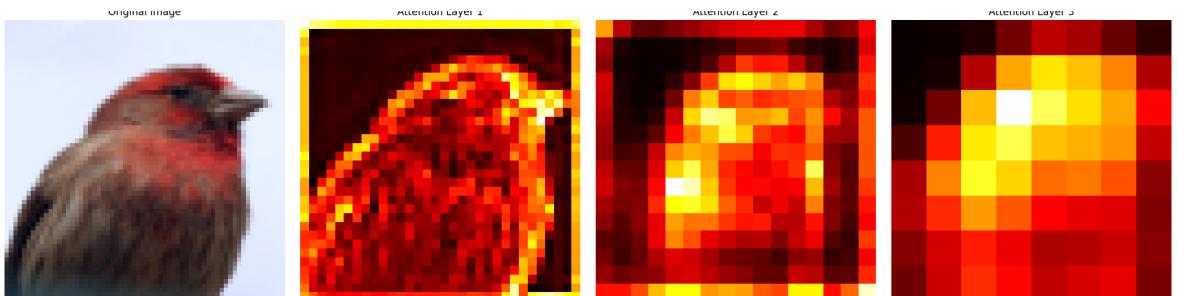


Рис. 4.2: Attention для Mini-ImageNet

4.3 Модель для DomainNet и Dogs vs. Cats

Далее, для исследования обобщающей способности CNN с inductive bias на Mini-ImageNet и оценки её универсальности, мы будем использовать датасеты DomainNet и Dogs vs Cats.

5 Обучение

В качестве функции потерь мы взяли кросс-энтропию. Это одна из самых распространённых функций потерь для задач многоклассовой классификации. В качестве оптимизатора мы взяли AdamW (Adam с возможностью добавить weight decay для регуляризации).

```
optim.AdamW(model.parameters(), lr=1e-4, weight_decay=1e-5) (1)
```

В качестве функции потерь используется кросс-энтропия с label smoothing 0.1:

```
nn.CrossEntropyLoss(label_smoothing=0.1) (2)
```

Также мы использовали планировщик Reduce on Plateau, который позволяет понижать learning rate при выходе значений функции потерь на плато.

Для того, чтобы добиться лучшего результата мы провели подбор гиперпараметров. Всего мы делали от 8 до 18 экспериментов, во время которых мы сравнивали комбинации значений различных значений гиперпараметров.

Гиперпараметр	Значения
learning rate	0.01, 0.001, 0.0001
batch size	64, 128
weight decay	0.0001, 0.001

Таблица 5.1: Гиперпараметры для датасета Food-101 (CNN)

Гиперпараметр	Значения
learning rate	0.001, 0.0001
batch size	64, 128
weight decay	0.0001, 0.001

Таблица 5.2: Гиперпараметры для датасета Food-101 (CNN + IB)

Гиперпараметр	Значения
learning rate	0.01, 0.001, 0.0001
batch size	32, 64, 128
weight decay	0.0001, 0.001

Таблица 5.3: Гиперпараметры для датасета Mini-ImageNet

Чтобы убедиться в стабильности обучения, были проведены эксперименты с разными random seeds для лучших моделей, которые были выявлены по подбору предыдущих параметров.

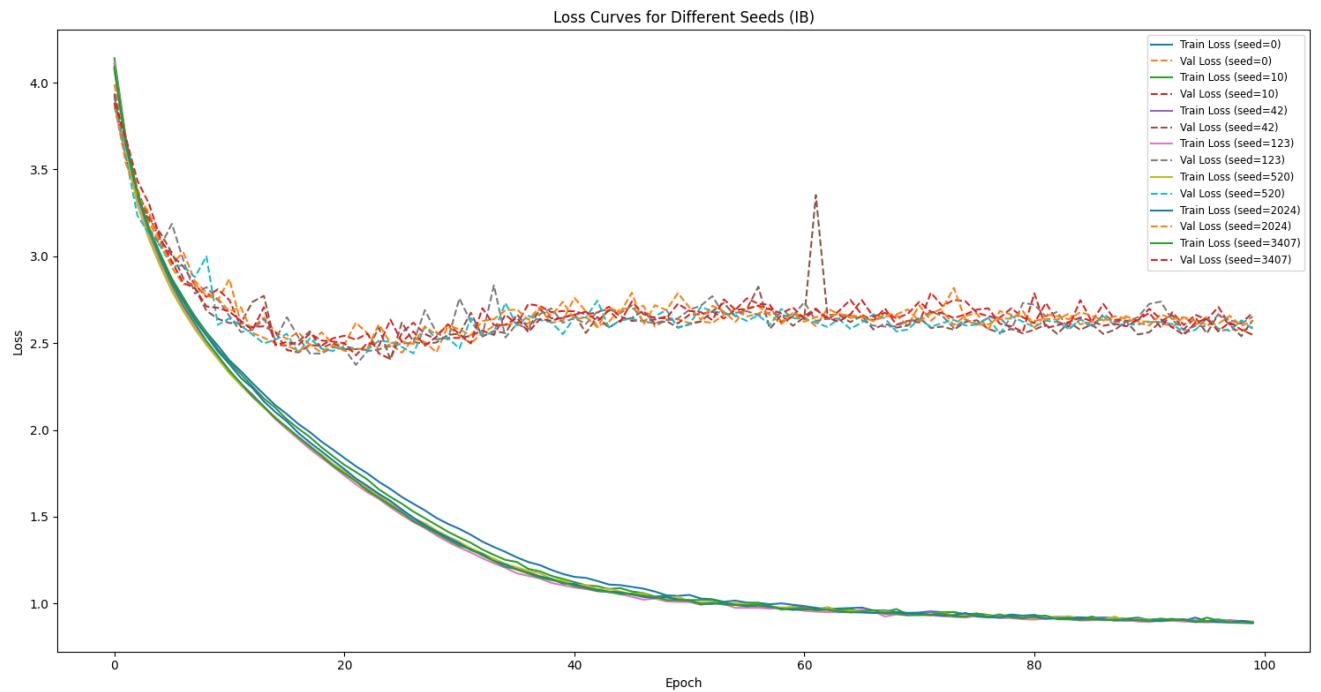


Рис. 5.1: Разные seed для CNN с Inductive Bias для Mini-ImageNet,

6 Результаты

6.1 Food-101

6.1.1 До внедрения Inductive Bias

Как и предполагалось, базовая CNN показывало достаточно низкое качество для всех экспериментов. Показатели функции потерь на валидации не опускалось ниже 3, что конечно же не самый хороший результат относительно других сетей. Однако, это довольно таки естественно, так как эта архитектура не очень глубокая.

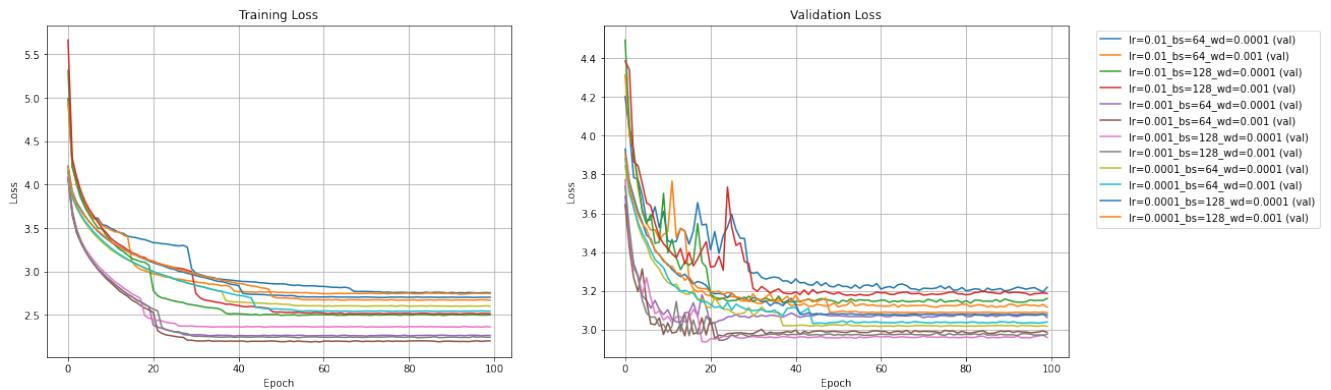


Рис. 6.1: Графики потерь для CNN без IB

Метрики качества тоже выдают низкие показатели: значение accuracy на валидации не превышает 0.32.

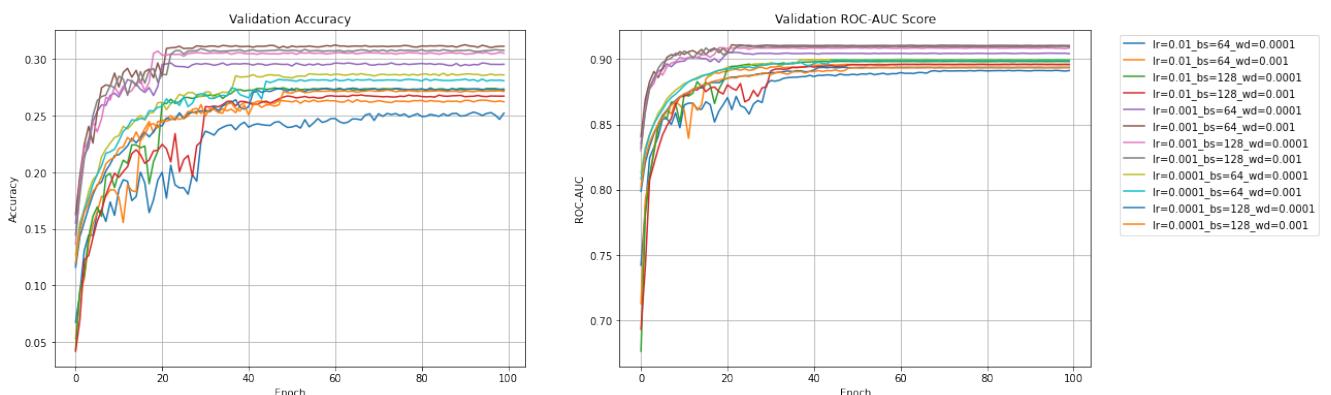


Рис. 6.2: Метрики для CNN без IB

6.1.2 После внедрения Inductive Bias

После внедрения Inductive Bias показатели функции потерь достаточно сильно улучшились, упав до значений 2.2 на валидации.

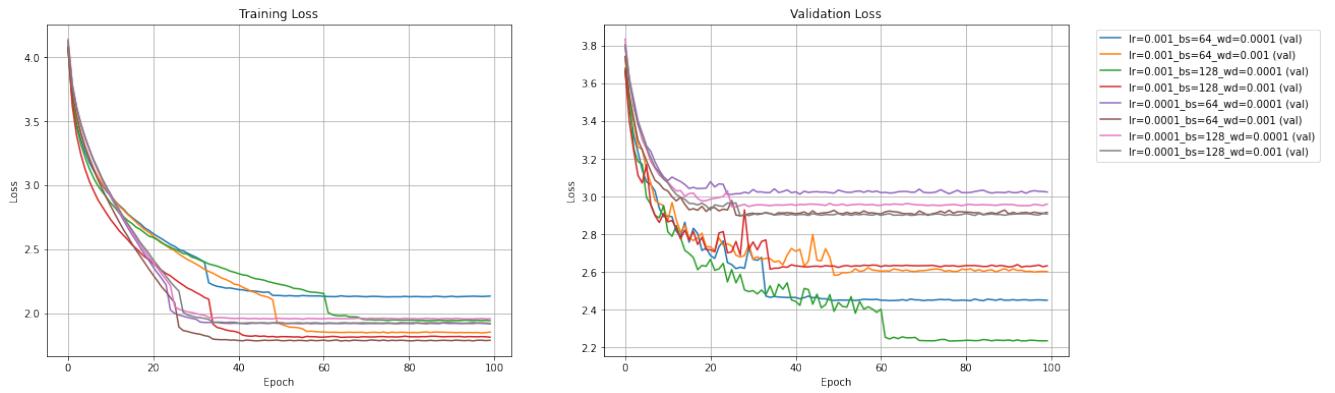


Рис. 6.3: Графики потерь для CNN с IB

Значения метрик качества тоже повысились: accuracy уже смогло добраться до значений 0.44-0.45, что превосходит предыдущие результаты примерно в 1.4 раза.

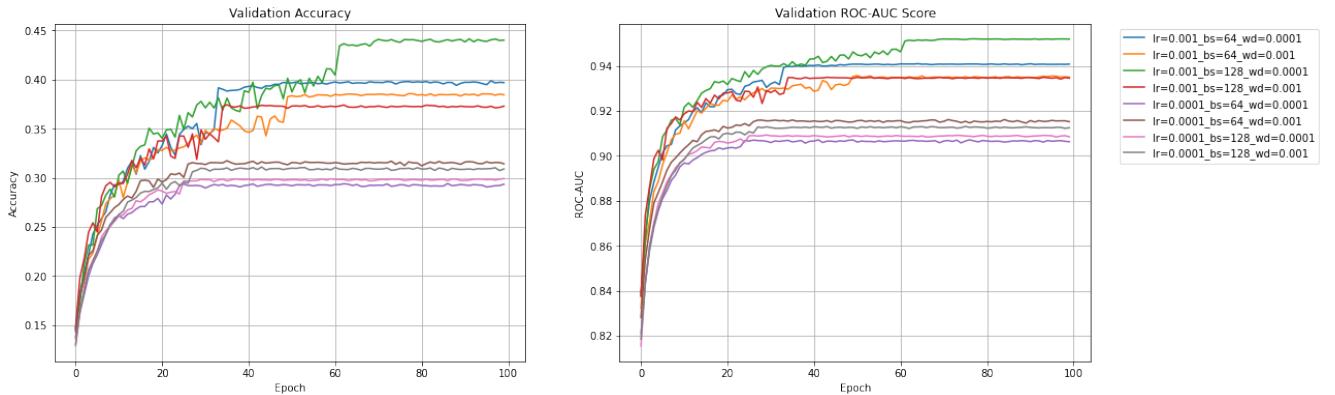


Рис. 6.4: Метрики для CNN с IB

6.1.3 Сравнение с другими бенчмарками

Датасет Food-101 очень часто используется для решения задачи классификации. Проводилось большое количество соревнований для решения этой задачи. И там мы сможем найти модели с качеством (accuracy) от 0.71 до 0.93 [2].

Конечно же наша модель не сравнима по качеству со всеми выше перечисленными. Однако, минимальное количество параметров у всех этих моделей примерно 110M (модель BERT + CNN [8]). В сравнении с нашей моделью с Inductive Bias, у которой количество параметров всего 9M, это весьма внушительная разница в размерах.

6.2 Mini-ImageNet

6.2.1 До внедрения Inductive Bias

Как и с предыдущим датасетом, качество обчной CNN оказалось очень низким (показатели функции потерь на валидации точно также не опускались ниже 3).

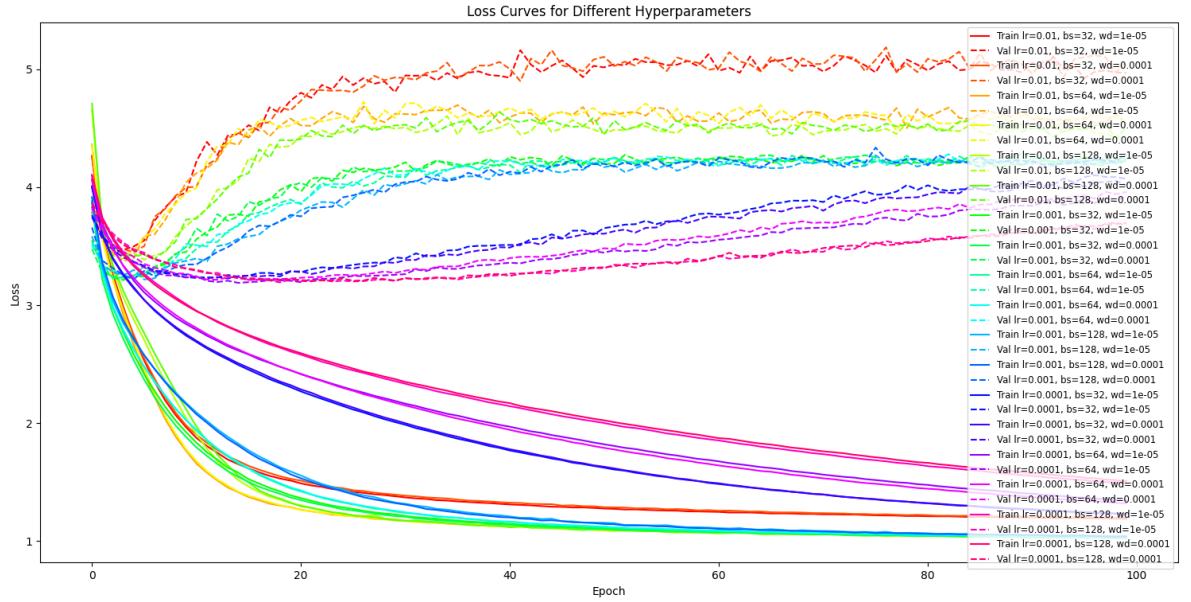


Рис. 6.5: Графики потерь для CNN без IB

Метрики качества тоже достаточно низкие: значение accuracy на валидации не превышает 0.35.

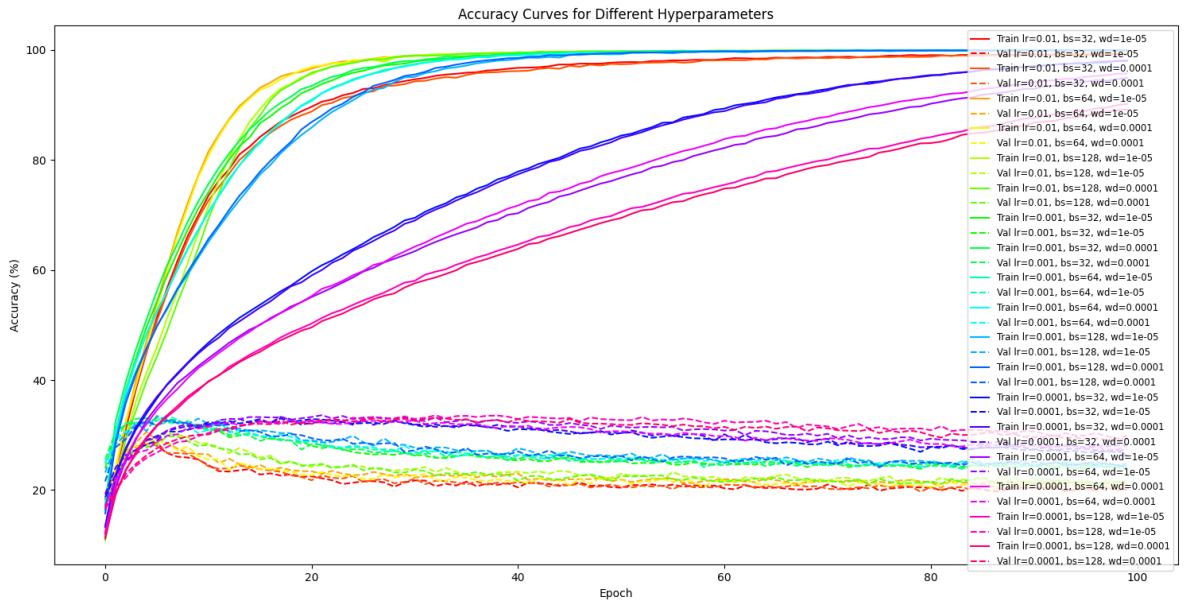


Рис. 6.6: Метрики для CNN без IB

6.2.2 После внедрения Inductive Bias

После внедрения Inductive Bias показатели функции потерь резко упали до значений 2.4 на валидации.

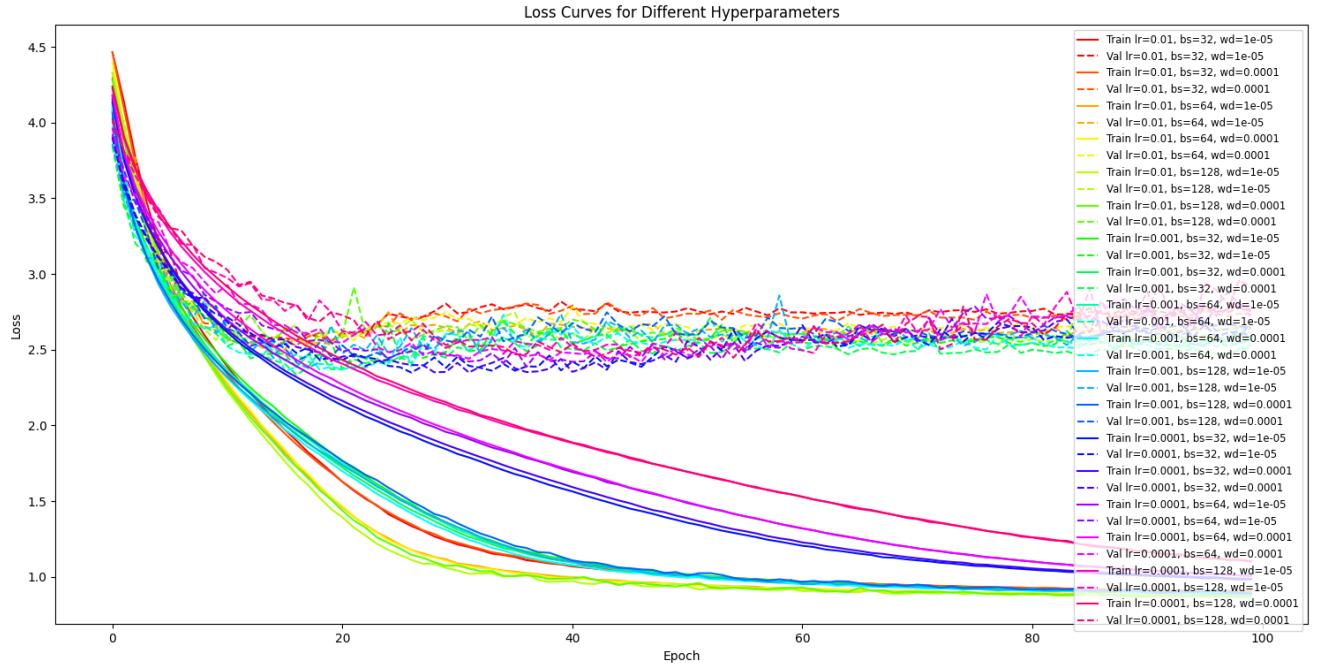


Рис. 6.7: Графики потерь для CNN с IB

Значения метрик качества сильно повысились: accuracy уже смогло добраться до значений 0.5-0.55. Это отличное повышение в качестве.

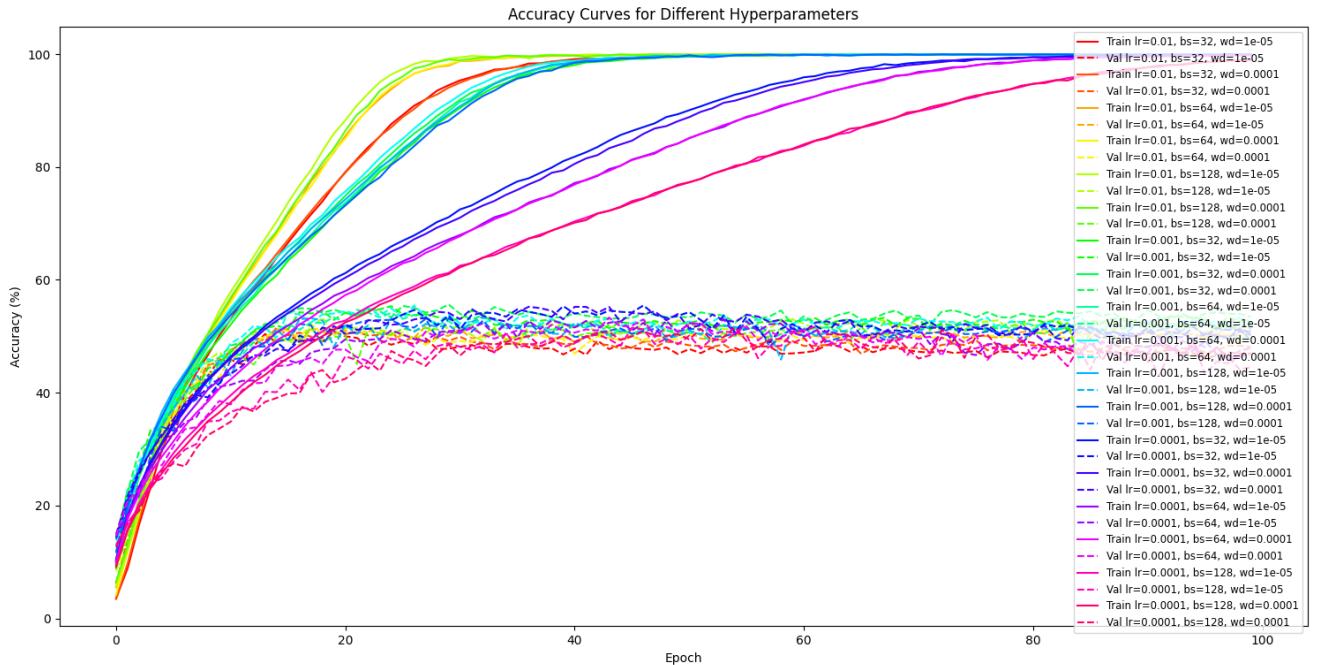


Рис. 6.8: Метрики для CNN с IB

6.2.3 Сравнение с другими бенчмарками

Модель с Inductive Bias выбрал baseline [3], несмотря на то что это простая CNN, но все равно получил хорошую accuracy — 53%. И после добавления Inductive Bias accuracy модели стал на 25% больше чем без Inductive Bias.

6.3 DomainNet

Далее мы решили проверить на обобщающую способность модели CNN with Inductive Bias для ImageNet: обучаем на одно стиле изображений, а классифицируем на другом.

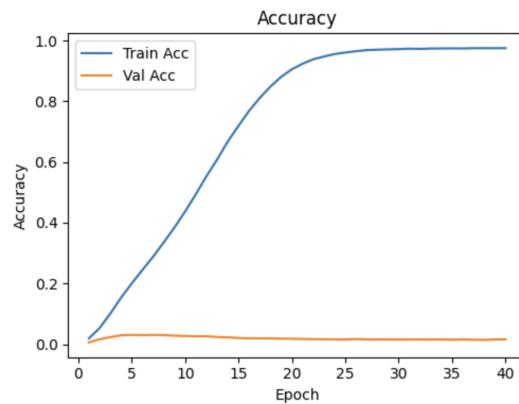


Рис. 6.9: DomainNet CNN with Infuctive Bias

Но к сожалению результат был очень плохим, так как в DomainNet 350 классов и у каждого класса только по 20-30 картинок. Что является очень сложной задачей для простой модели CNN. Тогда мы решили сделать бинарную классификацию на Dogs _ vs _ cats.

6.4 Dogs vs Cats

В данном датасете уже более качественные ресурсы, для train - по 500 реальных кошек и собак. А для val, test — ручные рисунки собак и кошек.

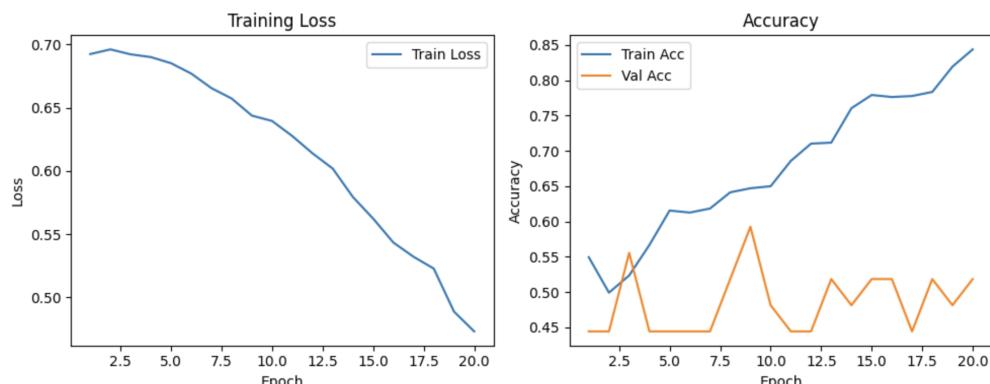


Рис. 6.10: Dogs vs Cats CNN with Infuctive Bias

Здесь уже видно, что мы получили хорошее качество, и значит модель имеет хорошую обобщающую способность.

7 Итоги

Правильное внедрение дополнительных признаков об устройстве данных внутрь модели иногда действительно позволяет улучшить качество во многих случаях. Более того, в некоторых случаях это способствует лучшему пониманию данных моделью, что повышает её обобщающую способность. Однако в то же время есть некий риск того, что обобщаемость снизится, так как, внедрив Inductive Bias, может получиться модель, подстроенная под какие-то конкретные данные.

Наша работа наглядно показывает пользу от выделения признаков из данных для внедрения их внутрь архитектуры моделей для задачи многоклассовой/бинарной классификации. Однако есть множество других сфер в машинном обучении, где внедрения этой информации могло бы повысить качество: от рекомендательных систем до генеративных моделей. Это открывает множество возможностей для дальнейшего исследования влияния Inductive Bias на нейронные сети, что делает нашу работу актуальной.

8 Обзор литературы

В статье **Inductive biases for deep learning of higher-level cognition** [9] рассказано про гипотезу, что человеческий и животный интеллект можно объяснить несколькими принципами, а не множеством эвристик, что упростило бы их понимание и создание ИИ. Изучение Inductive Bias, используемый людьми и животными, может помочь выявить эти принципы и вдохновить ИИ-исследования. Это особенно важно для развития ИИ, способного к гибкому обобщению и преодолению разрыва между машинным обучением и человеческим интеллектом.

В работе **Neural Production Systems** [1] рассматривается применение структурных индуктивных смещений для решения задач в визуальных средах с взаимодействующими объектами. В отличие от графовых нейронных сетей, которые зачастую неэффективно обрабатывают разреженные взаимодействия и не позволяют явно декомпозировать знания об объектах, авторы предлагают использовать производственные системы. В таких системах правила зависят от свойств объектов, что позволяет более гибко моделировать сложные взаимодействия. Такой подход, основанный на индуктивных смещениях, способствует улучшению предсказания будущих состояний системы и эффективному переносу знаний на более сложные среды.

Далее в статье **Towards Causal Representation Learning** [4] мы обсуждаем графовую причинность и машинное обучение, показывая, с каким IB из причинно-следственного анализа, можно решить главные проблемы машинного обучения, такие как перенос обучения и пр.; также в статье подробно разбирается обучение причинности (то есть выявление высокоуровневых причинных переменных из низкоуровневых данных), которое как правило объединяет обе области, делая новые индуктивные смещения не для графического анализа.

Список литературы

- [1] Anirudh Goyal Aniket Didolkar. *Neural Production Systems*. URL: <https://www.semanticscholar.org/paper/57fb3190887d837fca47a0ca176abc782b1f42d3> (дата обр. 23.03.2022).
- [2] *Baselines for Food-101*. URL: <https://paperswithcode.com/sota/image-classification-on-food-101-1>.
- [3] *Baselines for Mini-ImageNet*. URL: <https://paperswithcode.com/sota/few-shot-image-classification-on-mini-2>.
- [4] Francesco Locatello Bernhard Schölkopf. *Towards Causal Representation Learning*. URL: <https://www.semanticscholar.org/paper/8f566001453bc6be0a935bf69ffd90d9db3af32b> (дата обр. 22.02.2021).
- [5] *Dogs vs. Cats*. URL: <https://paperswithcode.com/dataset/cats-vs-dogs>.
- [6] *DomainNet*. URL: <https://paperswithcode.com/dataset/domainnet>.
- [7] *Food-101*. URL: <https://paperswithcode.com/dataset/food-101>.
- [8] Ignazio Gallo. *Image and Text fusion for UPMC Food-101 using BERT and CNNs*. URL: <https://artelab.dist.uninsubria.it/res/research/papers/2020/2020-IVCNZ-Gallo-Food101.pdf>.
- [9] Anirudh Goyal. *Higher-Level Cognitive Inductive Biases*. URL: <https://www.semanticscholar.org/paper/7e38476342ce1fcc8ef0dcd23686539395961769> (дата обр. 01.08.2022).
- [10] *Mini-ImageNet*. URL: <https://paperswithcode.com/dataset/mini-imagenet>.