

Регуляризация свёрточной нейронной сети сингулярным разложением для обучения на малых выборках

Н.В. Ваулин¹

Задача обучения на малой выборке решается за счет оптимизации параметров предобученной сверточной нейронной сети с ограничениями на веса сверток. Регуляризация весов осуществляется путем представления параметров свертки в виде набора сингулярных разложений матриц и обучения только сингулярных значений разложений. Исследуется влияние сингулярных значений на качество модели и динамика их изменений во время обучения. Приведены результаты применения предложенной регуляризации в сравнении с другими методами в задаче классификации рентгенологических снимков.

Ключевые слова: машинное обучение, свёрточные сети, разложение на сингулярные значения.

1. Введение

Автоматизированный анализ медицинских изображений является одной из наиболее перспективной областью применения машинного обучения. Задачу классификации медицинских изображений можно рассматривать в контексте общей проблемы классификации в компьютерном зрении. Здесь так же применимы многие стандартные решения, такие как нейросетевые архитектуры, аугментация данных и стратегии обучения. Однако, отсутствие больших обучающих выборок с одновременной необходимостью высокой обобщающей способности модели, заставляет развивать ряд специфичных решений, апеллирующих к недостатку данных. К ним относится дистилляция, перенос знаний, доменная адаптация – данные методы позволяют обучать качественные модели на малых выборках. Также при решении подобных задач используются нейросетевые архитектуры с малым числом параметров и блоки внимания.

Наиболее распространенный подход к обучению нейронных сетей на малых выборках является использование предобученных моделей. В данном случае возникает вопрос выбора базы для предобучения. Одним из

¹*Ваулин Николай Владимирович* — аспирант каф. теоретической информатики и дискретной математики, институт математики и информатики МПГУ, e-mail: nvvaulin@gmail.com.

Vaulin Nikolay Vladimirovich — graduate student, MPGU, Faculty of Computer Science and Discrete Mathematics.

вариантов такой базы является ImageNet[14] , что было опробовано в [4]. Однако, отличия в задачах классификации объектов и классификации болезней осложняют процесс адаптации предобученной модели к основной задаче. Таким образом, в [11] было продемонстрировано, что дальнейшее дообучение модели на более близком домене позволяет получить более информативное пространство признаков. Для этих целей авторы статьи собрали выборку из 224,316 КТ снимков грудной клетки с 14 размеченными патологиями на них. Бодель, предобученную на схожей выборке, можно адаптировать для решения нужной задачи. В [9] обучали последний классификационный слой для адаптации модели к задаче детекции COVID-19. Аналогичную задачу в статье [5] было предложено решить за счет обучения классификатора методом опорных векторов на выходе предпоследнего слоя. В статье [8] также решалась задача классификации для выявления COVID-19 на снимках грудной клетки. В данном случае, авторы статьи использовали densenet [12], предобученную на cheexpert [11], обучили последний классификационный слой и затем всю модель целиком.

Помимо предобучения модели на размеченной выборке, существуют методы, позволяющие задействовать в обучении неразмеченные данные. В [7] обучалась энкодер-декодер модель на снимках человеческого мозга, и затем энкодер использовался для выделения признаков, на которых обучался полносвязный слой в сценарии классификации, используя небольшую выборку. В работе [6] авторы использовали неразмеченную выборку, присвоив разметку на основе названия изображения, что во многом не соответствовало каким-либо патологиям, но хорошо подходило для предобучения модели.

Помимо расширения выборки, эффект переобучения снижают за счет уменьшения количества параметров модели. В статье [3] была предложена архитектура Covid-Net, в которой для снижения количества обучаемых параметров широко использовались PERX блоки, в которых сначала происходило проецирование признаков на меньшую размерность за счет свертки 1×1 , затем размерность увеличивалась сверткой 1×1 и применялась поканальная свертка 3×3 , далее размерность снова снижалась и восстанавливалась входная размерность для residual суммирования. Данная архитектура значительно снижала количество весов сверток. В работе [2] было показано, что архитектура из 4-х сверток 32-64-64-128 может превзойти в качестве задачи классификации COVID-19 глубокие архитектуры, такие как Resnet-50 и VGG-19, предобученные на imagenet.

Сингулярное разложение уже широко применяется в нейронных сетях. Например, в задаче распознавания речи удалось снизить количество параметров нейронной сети за счет обнуления малых сингулярных значений [16], уменьшив количество параметров в 8 раз и дообучив, точ-

ность модели практически не снизилась. Аналогично, снижение ранга матрицы использовалось в задаче дообучения (а именно, в self-supervised distillation). В [17] было предложено дистиллировать корреляцию между входными и выходными значениями блока при этом, внутренняя размерность входных и выходных признаков снижалась за счет обнуления сингулярных значений. Дообучение в режиме дистилляции к изначальной модели привело к увеличению точности.

В данной работе предлагается использовать сингулярное разложение весов в задаче переноса знаний при адаптации модели к новой задаче. За счет обучения только сингулярных значений сверточных матриц удастся добиться значительного уменьшения количества обучаемых параметров, что позволяет адаптировать модель для конкретной задачи, снизив эффект переобучения.

2. Перенос знаний

Рассмотрим задачу многоклассовой классификации в общем случае. Задача обучения сводится к максимизации правдоподобия параметров модели. Имея достаточно большой размер выборки, задачу можно решить минимизируя эмпирическое математическое ожидание логарифма правдоподобия, что реализуется алгоритмами стохастического градиентного спуска. Однако, в случае, когда количество обучаемых параметров сильно превышает размер обучающей выборки, оценка математического ожидания эмпирическим средним может давать смещенный результат, что проявляется в виде переобучения модели, оптимизация на маленькой выборке в конечном итоге может привести к росту логарифма правдоподобия, что обычно видно на валидационной базе. В данном сценарии широко применяются дополнительные ограничения на обучаемые параметры. Классическим предположением является близость оптимальных весов для схожих задач. Например, имея похожую выборку с большим количеством примеров, модель сначала обучается на ней, затем, осуществляется тонкая настройка параметров, за счет обучения на целевой задаче, используя малый шаг градиентного спуска. В данном случае, веса модели оказываются близки в евклидовой метрике к предобученным весам, что снижает эффект переобучения.

2.1. Сохранение базиса сингулярных векторов

Понятие близости весов можно сформулировать иначе. Рассмотрим веса свертки, их можно представить в виде матрицы (набора матриц). Каждую матрицу свертки исходной модели можно представить в виде сингулярного разложения $M = U\Sigma V^T$. При этом, можно сделать допу-

щение, что при дообучении меняются лишь сингулярные значения, при том что сингулярные вектора сверток остаются неизменными. Данный эффект выражен и без каких-либо дополнительных изменений во время обучения. Чтобы это продемонстрировать, применим матрицы U, V разложения исходных весов к дообученным $\Sigma_{tune} = U^T M_{tune} V$ и к весам модели, обученной со случайной инициализацией $\Sigma_{rand} = U^T M_{rand} V$. Затем сравним соотношение суммы квадратов диагональных элементов к сумме квадратов недиагональных $R = \sum_i \Sigma_{ii}^2 / \sum_{i \neq j} \Sigma_{ij}^2$. На рис. 1 изображено распределение этого соотношения. Среднее значение для модели, обученной со случайной инициализацией $\overline{R_{rand}} = 0.008$, в то время как для дообученной модели аналогичный параметр более чем в 100 раз выше $\overline{R_{tune}} = 0.91$. Данный эксперимент показывает состоятельность ограничения на дообучаемые параметры в виде фиксации матриц U и V сингулярного разложения сверток.

При вышеописанных ограничениях на веса, близость параметров выражается в схожести сингулярных значений. Забегая вперед, можно отметить, что данное предположение хорошо выполняется: корреляция между исходными сингулярными значениями и значениями дообученной модели при фиксированных U и V составляет 0.9.

2.2. SVD разложение свертки

Веса свертки нейронной сети представляют собой четырехмерный тензор $W \in R^{in, out, h, w}$. Один из способов разложения предполагает объединение размерности, соответствующей выходным каналам, и пространственных размерностей $M \in R^{in, out \cdot h \cdot w}$. Применив разложение на сингулярные значения, матрицу можно представить в виде $W = U \Sigma V^T$, где $U \in R^{out, out}$ и $V \in R^{out \cdot h \cdot w, out \cdot h \cdot w}$ - унитарные матрицы, а Σ - диагональная, с неотрицательными элементами. Зафиксировав матрицы U и V , можно уменьшить количество обучаемых параметров с $in \cdot out \cdot h \cdot w$ до $\min(in, out \cdot h \cdot w)$. Данный тип разложения будем называть полным.

Другой способ предполагает представление W в виде набора матриц $W_{ij} \in R^{in, out}$. В данном случае, сингулярное разложение применяется к каждой матрице в отдельности, при этом количество обучаемых параметров будет $\min(in, out) \cdot h \cdot w$, данный тип разложения будем называть пространственным, т.к. каждой пространственной компоненте свертки соответствует свое разложение.

Для еще большего снижения количества параметров, можно занулить малые сингулярные значения.

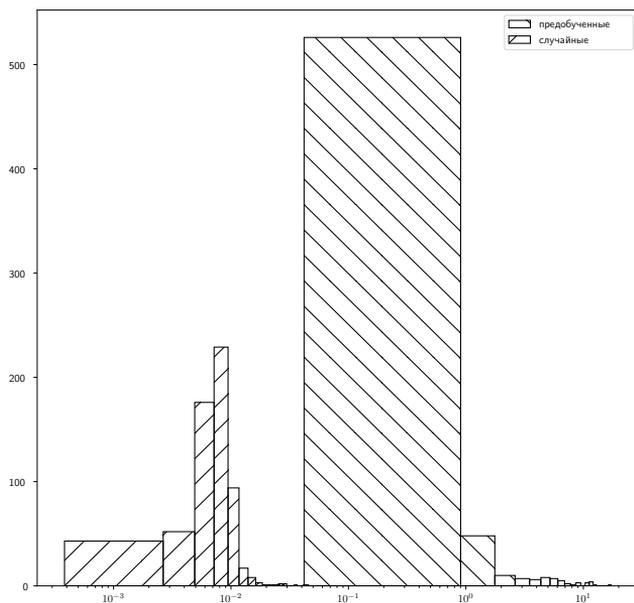


Рис. 1. Гистограмма отношения суммы квадратов диагональных элементов матрицы Σ к недиагональным в разложении по матрицам поворота исходных параметров для дообученной модели и обученной со случайной инициализации.

3. Постановка эксперимента

3.1. Данные

В качестве исходного датасета для предобучения модели была выбрана база NIH Chest X-ray[10]. Датасет содержит 112120 фронтальных рентгеновских изображений 30805 уникальных пациентов с размеченными на них 14 патологиями (каждое изображение может иметь несколько меток), полученных из радиологических отчетов с использованием автоматизированной обработки языка. Четырнадцать патологий грудной клетки включают ателектаз, уплотнение, инфильтрацию, пневмоторакс, отек, эмфизему, фиброз, выпот, пневмонию, уплотнение плевры, кардио-мегалию, узелок, объем и грыжу. Стоит отметить, что оригинальные радиологические отчеты не доступны для исследователя, лишь метки, точность которых, как утверждают авторы, составляет не менее 90%.

В качестве целевой задачи была выбрана задача многоклассовой классификации фронтальных рентгенологических снимков грудной клетки на 3 класса: COVID-19 (SARS-CoV-2), пневмония и без патологий. Для данных целей была выбрана обучающая выборка CovidAID[1], состоящая из 478 изображений с COVID-19 взятых из базы [18], 1583 снимков здоровых людей и 4273 снимков людей с пневмонией взятых из [19]. Для всех снимков есть уникальный идентификатор пациента, а точность разметки данных значительно выше, чем в NIH Chest X-ray[10], т.к. использовалась ручная разметка.

Все снимки имеют высокое разрешение, однако, входное разрешение обучаемой модели - квадрат 224 на 224 пиксела, соответственно, разрешение изображений снижалось до входного разрешения модели. В качестве аугментации использовались случайное отражение относительно вертикальной оси и случайная рамка с изменением масштаба в диапазоне от 0.8 до 1.2.

Вся база была разбита на 3 части: обучающая выборка более 88%, валидационная менее 2% и тестовая 10% (на Рис. 2 приведены подробная статистика базы и примеры изображений). При этом, все снимки одного пациента полностью включены в одну из трех подвыборок.

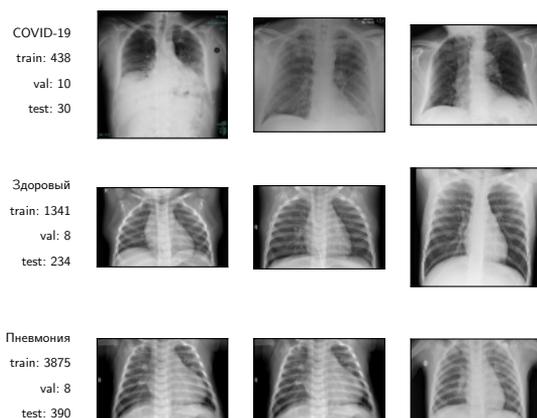


Рис. 2. Примеры изображений и статистика базы CovidAID.

3.2. Модель обучения

В качестве классификатора была выбрана архитектура нейронной сети `densenet`[12] с размером входного изображения 224 пикселей. В статье [13] была продемонстрирована высокая эффективность модели на базе NIH Chest X-ray[10]. Сеть изначально была обучена на `imageNet`[14], затем был заменен последний классификационный слой и сеть была дообучена на NIH Chest X-ray с помощью `Adam`[15] оптимизатора с мультипликатором градиентов 0.001. Полученные веса использовались во всех экспериментах в качестве инициализации.

Так как количество классов в целевой задаче отличалось от исходной, модель обучалась в 2 этапа: на первом этапе, модель, за исключением полносвязного слоя, инициализировалась предобученными весами и обучался только последний полносвязный слой одну эпоху, затем обучалась вся модель в течении 14 эпох. Во всех экспериментах был использован оптимизатор `Adam`[15] с $\beta_1 = 0.9, \beta_2 = 0.99$, размером батча 16 и мультипликатором градиентов 10^{-3} , причем мультипликатор снижался, если значения функции потерь на валидации не изменялось в течении четырех эпох. Лучшая эпоха выбиралась на основе значения функции потерь на валидационной выборке.

Так как сингулярные значения матрицы неотрицательные, во время обучения они были представлены в виде квадрата обучаемых параметров (соответственно, инициализировались эти параметры квадратными корнями сингулярных значений).

Для сравнения эффективности предложенного метода, было проведено 5 серий экспериментов.

- Обучение всех параметров весов, как было предложено авторами статьи [1]. Стоит отметить, что параметры обучения не соответствуют референсным, однако использовалась та же база, а результаты на тесте оказались выше, чем было продемонстрировано в статье (roc auc 0.9897 против 0.986), что позволяет использовать данный подход в качестве базового метода для сравнительного анализа (обозначено как **full**).
- Обучение только последнего классификационного слоя по аналогии со статьей [5] (обозначено как **fc-train**).
- Обучение сингулярных значений при полном разложении (обозначено как **full-svd**).
- 2 серии обучения сингулярных значений при пространственном разложении и занулении 0% и 90% сингулярных значений (обозначено как **sp-svd** и **sp-svd9**).

3.3. Тест и мера качества

Для сравнения результатов с базовыми методами была использована подвыборка CovidAID [1]. Мерой качества было выбрано значение площади под ROC-кривой для каждого класса (ROC AUC). Также измерялось отношение верноположительных детекций к общему количеству положительных примеров при пороге, соответствующему доле ложноположительных ответов равной 10^{-3} ($\text{tpr} @ \text{fmr}=10^{-3}$). Использование данной меры оправдано для измерения качества детектирования COVID-19, так как в тесте соответствующих примеров значительно меньше, чем примеров с пневмонией или здоровых людей. В таблице 1 приведены результаты для каждой серии экспериментов. Видно, что все 3 метода регуляризация svd разложением в среднем превосходят базовые методы, лишь в некоторых случаях уступая в детектировании COVID-19.

Таблица 1. Сравнение качества на тесте CovidAID

roc auc					
	fc-train	full	full-svd	sp-svd	sp-svd-9
Здоровый	0.9790	0.9837	0.9882	0.9928	0.9901
COVID-19	0.9952	1.0000	0.9996	0.9999	1.0000
Пневмония	0.9663	0.9855	0.9899	0.9933	0.9906
Среднее	0.9802	0.9897	0.9926	0.9953	0.9936
tpr @ fmr= 10^{-3}					
	fc-train	full	full-svd	sp-svd	sp-svd-9
Здоровый	0.2265	0.5641	0.4829	0.6667	0.4530
COVID-19	0.5667	1.0000	0.9000	0.9333	1.0000
Пневмония	0.1667	0.3974	0.3641	0.5487	0.4462
Среднее	0.3199	0.6538	0.5823	0.7162	0.6330

Стоит отметить, что в данном тесте всего 30 примеров с COVID-19 и меры качества достигали своих максимальных значений в некоторых экспериментах, что свидетельствует о недостаточном размере тестовой подвыборки. Также малый размер валидационной базы мог привести к неоптимальному выбору лучшей эпохи. В целях расширения тестовой выборки разбиение базы на обучающую, тестовую и валидационную было изменено: к исходному тесту и валидации были добавлены примеры за счет сокращения базы обучения, таким образом база была разделена в соотношении 50%, 40%, 10% (обучающая, тестовая и валидационная выборки соответственно), подробная статистика разбиения приведена в табл. 2. Все эксперименты были проведены повторно на новой обучаю-

щей выборке, результаты приведены в табл. 3. Данные результаты во многом повторяют предыдущие, однако, заметны отличия в качестве детектирования COVID-19.

Таблица 2. Статистика обучающей выборки

CovidAID				
	COVID-19	Здоровый	Пневмония	Сумма
Обучение	438	1341	3875	5654
Валидация	10	8	8	26
Тест	30	234	390	654
Сумма	478	1583	4273	6334
Измененный CovidAID				
	COVID-19	Здоровый	Пневмония	Сумма
Обучение	244	757	2168	3169
Валидация	49	152	431	632
Тест	185	674	1674	2533
Сумма	478	1583	4273	6334

Таблица 3. Сравнение качества на расширенном тесте CovidAID

roc auc					
	fc-train	full	full-svd	sp-svd	sp-svd-9
Здоровый	0.9841	0.9908	0.9943	0.9943	0.9928
COVID-19	0.9903	0.9998	0.9999	0.9998	0.9998
Пневмония	0.9696	0.9908	0.9945	0.9949	0.9938
Среднее	0.9813	0.9938	0.9962	0.9964	0.9955
tpr @ fmr=10 ⁻³					
	fc-train	full	full-svd	sp-svd	sp-svd-9
Здоровый	0.4525	0.5786	0.6736	0.6751	0.5445
COVID-19	0.3514	0.9622	0.9892	0.9405	0.9838
Пневмония	0.1147	0.2987	0.3363	0.5299	0.4217
Среднее	0.3062	0.6132	0.6664	0.7152	0.6500

3.4. Связь качества обучения и размера выборки

Для определения зависимости значения мер качества от размера обучающей выборки, были проведены эксперименты на базе, уменьшенной

в 2,4 и 8 раз. Для всех серий были использованы одинаковые подвыборки обучающей базы. В соответствии с результатами, приведенными в табл. 4, предложенная регуляризация показывает значительные преимущества при уменьшении базы. Однако, по графику изображенному на рис. 3 видно, что при уменьшении базы полное разложение показывает лучшие результаты, чем пространственное. В то же время, разрыв между обучением всех сингулярных значений и отброс 90% значений, снижается при уменьшении базы. Таким образом заметна зависимость качества обучения от количества обучаемых параметров, полное разложение имеет сильно меньше обучаемых параметров, и показывает себя лучше при уменьшении базы, в то время как качество при обучении всех параметров резко падает.

Таблица 4. Сравнение качества модели при уменьшении обучающей базы на расширенном тесте CovidAID

roc auc					
Размер подвыборки	fc-train	full	full-svd	sp-svd	sp-svd-9
1	0.9813	0.9938	0.9962	0.9964	0.9955
1/2	0.9771	0.9889	0.9952	0.9958	0.9937
1/4	0.9762	0.9904	0.9937	0.9941	0.9928
1/8	0.9701	0.9645	0.9916	0.9863	0.9907
tpr @ fmr=10 ⁻³					
Размер подвыборки	fc-train	full	full-svd	sp-svd	sp-svd-9
1	0.3062	0.6132	0.6664	0.7152	0.6500
1/2	0.2244	0.5238	0.6635	0.6740	0.5716
1/4	0.1958	0.4336	0.4928	0.5674	0.5373
1/8	0.2395	0.2544	0.4730	0.3942	0.4721

4. Анализ результатов

Предложенный подход позволяет регуляризовать параметры сети и избежать эффекта переобучения даже на небольшой обучающей выборке, в то же время, оптимальный метод разложения весов зависит от размера выборки. Метод позволяет гибко уменьшать количество обучаемых параметров, если это требуется при имеющихся данных для обучения. Также метод позволяет уменьшить общее количество параметров сети: при отбрасывании 90% параметров сети, качество модели снижалось незначительно, таким образом можно изменить архитектуру сети в целях снижения вычислительной сложности или для сжатия параметров

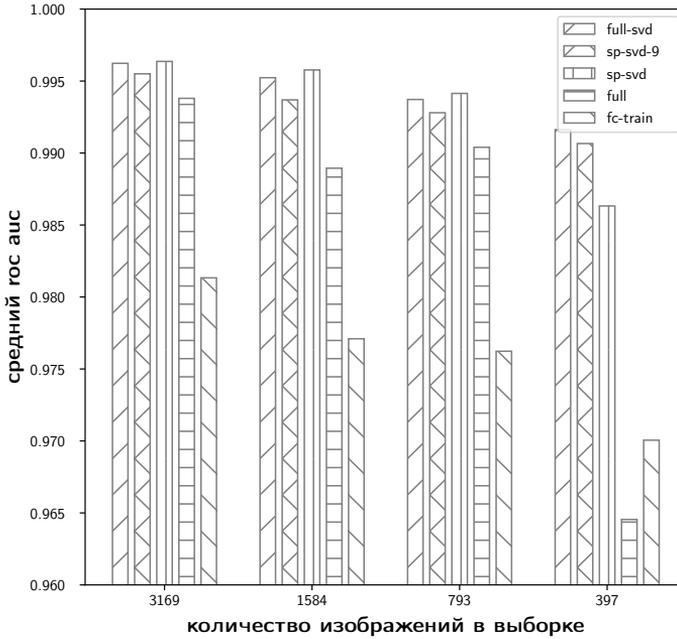


Рис. 3. Зависимость качества обучения (гос аус на расширенном CovidAID) при уменьшении базы в 1, 2, 4 и 8 раз.

сети. Данный результат во многом повторяет [16], но уже в применении к сверточным сетям.

4.1. Динамика сингулярных значений

Рассмотрим взаимосвязь между исходными и дообученными сингулярными значениями. Так как после свертки всегда используется батч нормализация, стоит нормализовать значения, в данном случае, все сингулярные значения свертки делились на максимальное значение. На рис. 4 приведены гистограммы сингулярных значений и их изменений. Видно, что малые значения еще больше уменьшились, что свидетельствует о вырождении шумовых компонент, при этом усиливались лишь значимые компоненты, нормализованные значения которых больше 0.4. Любопытным является тот факт, что более половины (55%) сингулярных значений занулились, при том, что изначально околонулевых значений было менее 1%.

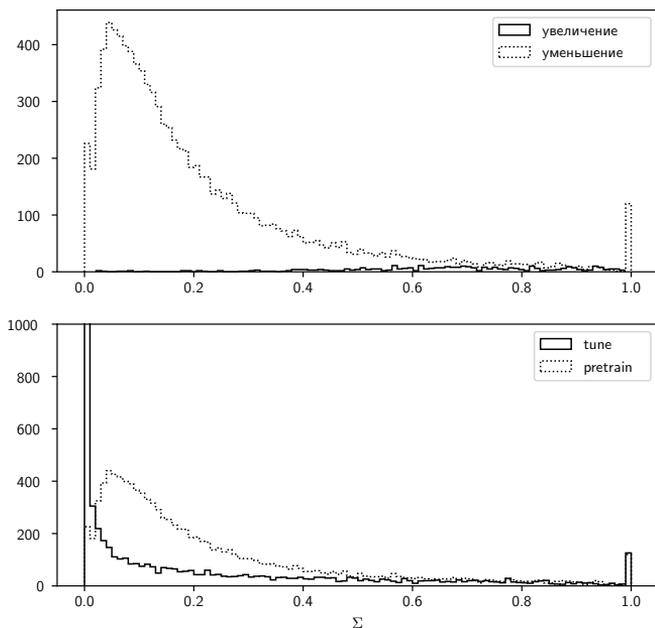


Рис. 4. Динамика изменений (сверху) и распределение (снизу) сингулярных значений предобученной модели (pretrain) и модели, дообученной на целевой задаче (tune).

4.2. Дисперсия обучения

Дополнительным преимуществом метода является его стабильность во время обучения. На рис. 5 видно, что метод полного разложения дает значительно более стабильные результаты в сравнении с обучением всех параметров: дисперсия значений функции потерь на разных эпохах более чем в 3 раза ниже (0.015 против 0.047), однако наиболее стабилен остается метод обучения последнего полносвязного слоя (дисперсия 0.009). Меньший разброс значений означает меньшую зависимость качества модели от выбора валидационной базы, что актуально при малом количестве данных, ввиду сложности сбора репрезентативной валидационной выборки.

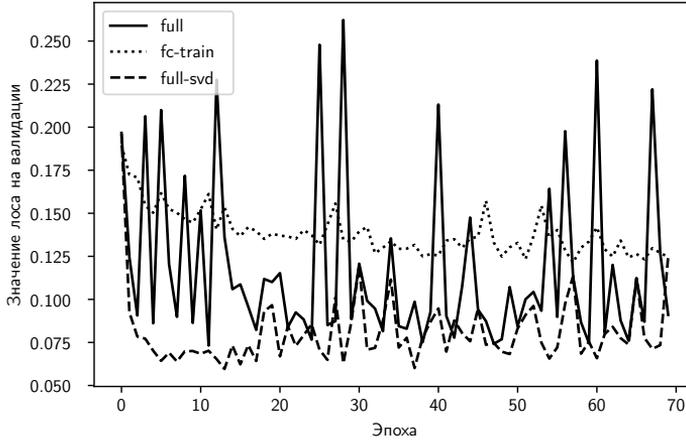


Рис. 5. Значения функции потерь на тестовой выборке в зависимости от эпохи.

5. Заключение

В работе предложен новый подход регуляризации параметров. Параметризация весов свертки в виде матриц сингулярного разложения и последующее дообучение лишь сингулярных значений ведет к снижению эффекта переобучения. Использование данного метода привело к повышению точности результатов на целевой задаче даже при кратном уменьшении размера обучающей выборки. Метод показал высокую стабильность во время обучения, что актуально в условиях отсутствия репрезентативной валидационной базы. Возможность уменьшения количества обучаемых параметров за счет зануления малых сингулярных значений позволяет эффективно использовать данный метод в широком диапазоне размеров обучающих выборок.

Побочным, но, несомненно, полезным, результатом метода является снижение ранга матриц свертки, что можно использовать для уменьшения количества параметров сети или для ее ускорения. Была продемонстрирована возможность эффективного сжатия сверточной сети счет обнуления сингулярных значений.

Список литературы

- [1] Mangal, Arpan and Kalia, Surya and Rajgopal, Harish and Rangarajan, Krithika and Namboodiri, Vinay and Banerjee, Subhashis and Arora,

- Chetan, “CovidAID: COVID-19 detection using chest X-ray”, *arXiv preprint arXiv:2004.09803*, 2020.
- [2] Haque, Khandaker Foysal and Abdelgawad, Ahmed, “A deep learning approach to detect COVID-19 patients from chest X-ray images”, *AI*, **1:3** (2020), 418–435.
- [3] Wang, Linda and Lin, Zhong Qiu and Wong, Alexander, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images”, *Scientific Reports*, **10:1** (2020), 1–12.
- [4] Raghu, Maithra and Zhang, Chiyuan and Kleinberg, Jon and Bengio, Samy, “Transfusion: Understanding transfer learning for medical imaging”, *arXiv preprint arXiv:1902.07208*, 2019.
- [5] Sethy, Prabira Kumar and Behera, Santi Kumari, “Detection of coronavirus disease (covid-19) based on deep features”, 2020.
- [6] Alzubaidi, Laith and Al-Amidie, Muthana and Al-Asadi, Ahmed and Humaidi, Amjad J and Al-Shamma, Omran and Fadhel, Mohammed A and Zhang, Jinglan and Santamarí a, J and Duan, Ye, “Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data”, *Cancers*, **13:7** (2021), 1590.
- [7] Chen, Min and Shi, Xiaobo and Zhang, Yin and Wu, Di and Guizani, Mohsen, “Deep features learning for medical image analysis with convolutional autoencoder neural network”, *IEEE Transactions on Big Data*, 2017.
- [8] Singh, Shrinjal and Sapra, Piyush and Garg, Aman and Vishwakarma, Dinesh Kumar, *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021.
- [9] Minaee, Shervin and Kafieh, Rahele and Sonka, Milan and Yazdani, Shakib and Soufi, Ghazaleh Jamalipour, “Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning”, *Medical image analysis*, **65** (2020), 101794.
- [10] Wang, X and Peng, Y and Lu, L and Lu, Z and Bagheri, M and Summers, R, *IEEE CVPR*, 2017.
- [11] Irvin, Jeremy and Rajpurkar, Pranav and Ko, Michael and Yu, Yifan and Ciurea-Ilcus, Silvana and Chute, Chris and Marklund, Henrik and Haghgoo, Behzad and Ball, Robyn and Shpanskaya, Katie and others, *Proceedings of the AAAI conference on artificial intelligence*, **33**, 2019.

- [12] Gao Huang and Zhuang Liu and Laurens van der Maaten and Kilian Q. Weinberger, “Densely Connected Convolutional Networks”, 2018.
- [13] Rajpurkar, Pranav and Irvin, Jeremy and Zhu, Kaylie and Yang, Brandon and Mehta, Hershel and Duan, Tony and Ding, Daisy and Bagul, Aarti and Langlotz, Curtis and Shpanskaya, Katie and others, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning”, *arXiv preprint arXiv:1711.05225*, 2017.
- [14] Deng, Jia and Dong, Wei and Socher, Richard and Li, Li-Jia and Li, Kai and Fei-Fei, Li, *2009 IEEE conference on computer vision and pattern recognition*, 2009.
- [15] Kingma, Diederik P and Ba, Jimmy, “Adam: A method for stochastic optimization”, *arXiv preprint arXiv:1412.6980*, 2014.
- [16] Xue, Jian and Li, Jinyu and Gong, Yifan, *Interspeech*, 2013.
- [17] Lee, Seung Hyun and Kim, Dae Ha and Song, Byung Cheol, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [18] Joseph Paul Cohen and Paul Morrison and Lan Dao and Karsten Roth and Tim Q Duong and Marzyeh Ghassemi, “COVID-19 Image Data Collection: Prospective Predictions Are the Future”, *arXiv 2006.11988*, 2020.
- [19] Kermany, Daniel S and Goldbaum, Michael and Cai, Wenjia and Valentim, Carolina CS and Liang, Huiying and Baxter, Sally L and McKeown, Alex and Yang, Ge and Wu, Xiaokang and Yan, Fangbing and others, “Identifying medical diagnoses and treatable diseases by image-based deep learning”, *Cell*, **172**:5 (2018), 1122–1131.

Neural network regularization via SVD for small dataset tasks

Vaulin N.V.

Consider convolutional neural network optimization problem using small dataset. The proposed method fine-tune pretrained neural network with specific constrains for convolutional kernels. Pretrained weights are decomposed using SVD. During fine-tuning the only singular values are trained. In the paper are investigated the dynamic of singular values during training process and its influence on quality of resulting model. The method are compared with other approaches in renegological images classification problem.

Keywords: deep learning, convolutional neural networks, singular value decomposition.

References

- [1] Mangal, Arpan and Kalia, Surya and Rajgopal, Harish and Rangarajan, Krithika and Namboodiri, Vinay and Banerjee, Subhashis and Arora, Chetan, “CovidAID: COVID-19 detection using chest X-ray”, *arXiv preprint arXiv:2004.09803*, 2020.
- [2] Haque, Khandaker Foysal and Abdelgawad, Ahmed, “A deep learning approach to detect COVID-19 patients from chest X-ray images”, *AI*, **1**:3 (2020), 418–435.
- [3] Wang, Linda and Lin, Zhong Qiu and Wong, Alexander, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images”, *Scientific Reports*, **10**:1 (2020), 1–12.
- [4] Raghu, Maithra and Zhang, Chiyuan and Kleinberg, Jon and Bengio, Samy, “Transfusion: Understanding transfer learning for medical imaging”, *arXiv preprint arXiv:1902.07208*, 2019.
- [5] Sethy, Prabira Kumar and Behera, Santi Kumari, “Detection of coronavirus disease (covid-19) based on deep features”, 2020.
- [6] Alzubaidi, Laith and Al-Amidie, Muthana and Al-Asadi, Ahmed and Humaidi, Amjad J and Al-Shamma, Omran and Fadhel, Mohammed A and Zhang, Jinglan and Santamarí a, J and Duan, Ye, “Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data”, *Cancers*, **13**:7 (2021), 1590.
- [7] Chen, Min and Shi, Xiaobo and Zhang, Yin and Wu, Di and Guizani, Mohsen, “Deep features learning for medical image analysis with convolutional autoencoder neural network”, *IEEE Transactions on Big Data*, 2017.
- [8] Singh, Shrinjal and Sapra, Piyush and Garg, Aman and Vishwakarma, Dinesh Kumar, *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021.
- [9] Minaee, Shervin and Kafieh, Rahele and Sonka, Milan and Yazdani, Shakib and Soufi, Ghazaleh Jamalipour, “Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning”, *Medical image analysis*, **65** (2020), 101794.
- [10] Wang, X and Peng, Y and Lu, L and Lu, Z and Bagheri, M and Summers, R, *IEEE CVPR*, 2017.

- [11] Irvin, Jeremy and Rajpurkar, Pranav and Ko, Michael and Yu, Yifan and Ciurea-Ilcus, Silvana and Chute, Chris and Marklund, Henrik and Haghighi, Behzad and Ball, Robyn and Shpanskaya, Katie and others, *Proceedings of the AAAI conference on artificial intelligence*, **33**, 2019.
- [12] Gao Huang and Zhuang Liu and Laurens van der Maaten and Kilian Q. Weinberger, “Densely Connected Convolutional Networks”, 2018.
- [13] Rajpurkar, Pranav and Irvin, Jeremy and Zhu, Kaylie and Yang, Brandon and Mehta, Hershel and Duan, Tony and Ding, Daisy and Bagul, Aarti and Langlotz, Curtis and Shpanskaya, Katie and others, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning”, *arXiv preprint arXiv:1711.05225*, 2017.
- [14] Deng, Jia and Dong, Wei and Socher, Richard and Li, Li-Jia and Li, Kai and Fei-Fei, Li, *2009 IEEE conference on computer vision and pattern recognition*, 2009.
- [15] Kingma, Diederik P and Ba, Jimmy, “Adam: A method for stochastic optimization”, *arXiv preprint arXiv:1412.6980*, 2014.
- [16] Xue, Jian and Li, Jinyu and Gong, Yifan, *Interspeech*, 2013.
- [17] Lee, Seung Hyun and Kim, Dae Ha and Song, Byung Cheol, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [18] Joseph Paul Cohen and Paul Morrison and Lan Dao and Karsten Roth and Tim Q Duong and Marzyeh Ghassemi, “COVID-19 Image Data Collection: Prospective Predictions Are the Future”, *arXiv 2006.11988*, 2020.
- [19] Kermany, Daniel S and Goldbaum, Michael and Cai, Wenjia and Valentim, Carolina CS and Liang, Huiying and Baxter, Sally L and McKeown, Alex and Yang, Ge and Wu, Xiaokang and Yan, Fangbing and others, “Identifying medical diagnoses and treatable diseases by image-based deep learning”, *Cell*, **172**:5 (2018), 1122–1131.