Послойная дистилляция знаний для обучения упрощенных биполярных морфологических нейронных сетей

М. В. Зингеренко [I, II], Е. Е. Лимонова [II, III]

[I] Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Московская область, Россия
[II] ООО «Смарт Энджинс Сервис», Москва, Россия
[III] Федеральный исследовательский центр «Информатика и Управление» РАН, Москва, Россия

Аннотация. В работе представлено улучшение структуры биполярного морфологического нейрона, повышающее его вычислительную эффективность и новый подход к обучению на основе непрерывных аппроксимаций максимума и дистилляции знаний. Были проведены эксперименты на выборке MNIST с нейронной сетью LeNet-подобной архитектуры, а также на выборке CIFAR10 с моделью архитектуры ResNet-22. На LeNet-подобной модели с помощью предложенного метода обучения получилось добиться 99.45% точности классификации при такой же точности у классической сети, а на ResNet-22 точность составила 86.69% при точности 86.43% у классической модели. Полученные результаты показывают, что предложенный метод, использующий log-sum-exp (LSE) аппроксимацию максимума и послойную дистилляцию знания, позволяет получить упрощенную биполярную морфологическую сеть, не уступающую классическим сетям.

Ключевые слова: биполярные морфологические сети, аппроксимации, искусственные нейронные сети, вычислительная эффективность.

DOI 10.14357/20718632230305

Введение

Современные системы распознавания образов трудно представить без нейронных сетей [1-3]. Нейронные сети сейчас активно используются на мобильных процессорах [4] и программируемых логических интегральных схемах [5]. Для улучшения производительности в таких системах активно разрабатываются различные подходы и методы, такие как, например, квантование нейронных сетей [6], тензорные декомпозиции [7] или удаление весов [8]. Одним из подходов является разработка специальных моделей нейронов, использующих более простые операции, чем классические модели, например, аддитивные сети [9-10]. Примером применения такого подхода является биполярный морфологический нейрон [11-12].

Биполярный морфологический (БМ) нейрон использует в своей основе операции сложения и взятия максимума, а не умножения и сложения, как классический математический нейрон. Операция сложения имеет меньшую аппаратную сложность, чем операция умножения, поэтому БМ нейрон потенциально более энергоэффективный и быстрый, чем классический математический нейрон. Однако основной проблемой

этой модели является трудность обучения градиентными методами на основе обратного распространения ошибки. Из-за наличия операции максимума во время обучения изменяются только 4 значения весов за одну итерацию для каждого нейрона. Также сама структура нейрона состоит из 4 вычислительных веток, то есть требует дополнительных ресурсов для реализации. В этой работе представлена новая структура БМ нейрона – 1.5-веточная модель, а также новый метод обучения для нее, позволяющий получить результаты не хуже, чем с помощью классической модели. Эксперименты показали, что на выборке MNIST при использовании LeNet-подобной архитектуры точность классификации предложенный 1.5-веточной БМ сети составила 99.45%, что не уступает точности классической сети. На выборке CIFAR10 1.5-веточная БМ сеть архитектуры ResNet-22 продемонстрировала точность распознавания 86.69%, в то время, как точность классической модели составила 86.43%.

1. Биполярный морфологический нейрон

Классическая модель нейрона может быть представлена следующим образом:

$$f(\vec{x}) = \phi\left(\sum_{i=1}^{n} \omega_i x^i - \omega_0\right),$$

где ω^+ и ω^- – векторы весов, x – вектор входных значений, а ϕ – функция активации. Такую модель можно приблизить с помощью биполярного морфологического нейрона, который записывается следующим образом:

$$f(\vec{x}) = \phi\{\exp\max_{i=1}^{n}(\ln x_i^+ + v_i^+) - \exp\max_{i=1}^{n}(\ln x_i^+ + v_i^-) - \exp\max_{i=1}^{n}(\ln x_i^- + v_i^+) + \exp\max_{i=1}^{n}(\ln x_i^- + v_i^-) + v_0\}, \quad (1)$$

$$x_i^+ = \begin{cases} 0, & x < 0, \\ x_i, & x \geq 0. \end{cases} \quad x_i^- = \begin{cases} -x_i, & x < 0, \\ 0, & x \geq 0. \end{cases}$$

где ω^+ и ω^- – векторы весов, x – вектор входных значений, а ϕ – функция активации. Как видно из выражения (1), такая структура имеет четыре вычислительных ветки, которые возникают из-за ограничений, связанных с областью определения операции взятия логарифма. Основной проблемой этой модели является трудность обучения стандартными методами: так как из-за наличия операции жесткого максимума (max) во время обучения в процессе обратного распространения ошибки изменяются только 4 значения весов за одну итерацию для каждого нейрона. Для решения этой проблемы в оригинальной работе был предложен метод послойного преобразования и дообучения.

2. Метод послойного преобразования и дообучения для БМ нейрона

Предложенный в оригинальной статье метод обучения заключается в послойном преобразовании нейронной сети к БМ виду и последующем дообучении модели стандартными методами (см. Алгоритм 1). На каждом шаге конвертировался один слой обученной классической сети к БМ виду, а далее выполнялось дообучение. Преобразование весов в одном слое проводилось следующим образом:

$$\omega_i^- = \begin{cases} \ln(|v_i|), & v_i < 0, \\ -\infty, & v_i \geq 0. \end{cases}$$
$$\omega_i^+ = \begin{cases} \ln(|v_i|), & v_i > 0, \\ -\infty, & v_i \leq 0. \end{cases} \quad (2)$$

где ω^+ и ω^- – векторы весов БМ сети, а v_i - вектор весов обученной классической сети.

Алгоритм 1 Обучение биполярной морфологической сети

1: *Обучить классическую сеть стандартным методом.*
2:
3: **Для каждого** *сверточного и линейного слоя* **выполнять**
4: *Произвести конвертацию по формуле (2) для текущего слоя.*
5: *Дообучить получившеюся сеть.*

3. Упрощенная 1.5-веточная модель БМ нейрона

Одним из недостатков БМ нейрона является его четырехветочная структура, поскольку она требует дополнительных аппаратных ресурсов для реализации. Однако если сместить входной вектор x и вектор весов ω в первую четверть, то надобность в наличии четырех вычислительных веток отпадет. Такую модель можно представить в следующем виде:

$$f(\vec{x}) = \phi\{\exp\max_{i=1}^{n}[\ln(x_i + \Delta x_i) + \ln(\omega_i + \Delta\omega_i)] - \Delta x_i\omega_i - \Delta\omega_i x_i - \Delta x_i\Delta\omega_i\}$$

где ω – вектор весов, x – вектор входных значений, ϕ – функция активации, а Δx и Δω – это смещения для входного вектора x и вектора весов ω, такие, чтобы значения под логарифмами были положительными. Назовем эту модель 1.5-веточным БМ нейроном, так как количество веток в данном варианте модели сокращается до одной, но добавляются дополнительные действия для обработки Δx и Δω. Однако предложенная модель не обучается с помощью классических методов обучения или методом, представленным выше, что можно увидеть в разделе 5, поэтому для нее был разработан метод обучения на основе дистилляции знания.

4. Обучение 1.5-веточных БМ сетей с помощью дистилляции знаний

4.1. Послойная дистилляции знаний

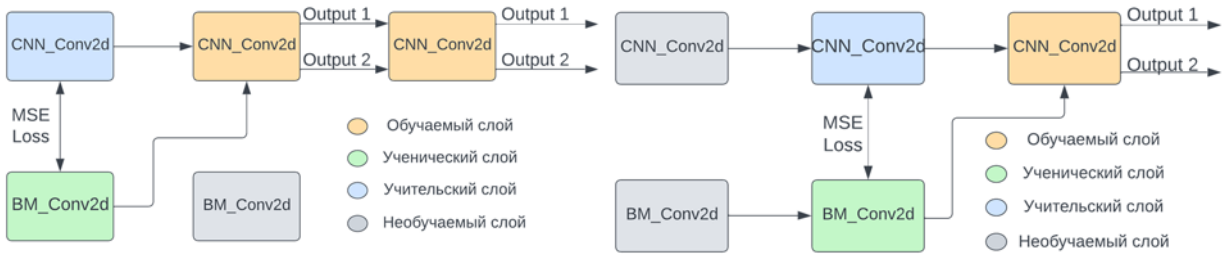
Суть обучения с помощью дистилляции знаний [13] состоит в передаче дополнительной полезной информации от учительской сети к ученической посредством изменения функции потерь следующим образом [14]:

$$L = \beta\sum_{i=1}^{n} H_{mse}(y_s^i, y_t^i) + \alpha\{H_{cross}(y_s, y_t) + H_{cross}(y_s, y_{gt})\}$$

где y_s^i – выход одного слоя ученической сети, y_t^i – выход одного слоя учительской сети, y_s , y_t – выходы всей ученической и учительской сети соответственно, y_{gt} – эталонный выход, n – число слоев в сети, H_{cross} – функция перекрёстной энтропии, H_{mse} – среднеквадратичная ошибка, α и β – температурные параметры. Таким образом нейронная сеть обучается не только на эталонных ответах, но и выход каждого слоя ученической сети становится похож на выход соответствующего слоя учительской сети. Так как БМ сеть является аппроксимацией классической сети, этот подход к обучению является оправданным.

4.2. Послойная дистилляция знаний для БМ сетей

Рассмотрим использование дистилляции знания для обучения БМ сетей. Принцип работы этого метода можно увидеть на Рис. 1. Во время обучения ученическая БМ сеть (BM) строится последовательно, слой за слоем. Для каждого добавляемого слоя к функции потерь добавляется среднеквадратичная ошибка между ним и классическим слоем учительской сети (CNN), после этого выход добавленного слоя передается в следующий слой учительской сети, где считаются два выхода $output_1$ и $output_2$ и общая функция потерь как сумма энтропии первого и второго выхода:



а) б)

Рис. 1. Принцип работы послойной дистилляции знания для БМ моделей для двух последовательно добавляемых слоев
а) для первого слоя, б) для второго слоя

$$L = \beta \sum_{i=1}^{n} H_{mse}(y_s^i, y_t^i) + \\ + \alpha\{H_{cross}(output_1, y_{gt}) + H_{cross}(output_2, y_{gt})\}$$

На следующем шаге только что обученный слой фиксируется и выполняется обучение следующего слоя. В итоге, к концу обучения будет обучена вся БМ сеть.

Этот подход позволяет нивелировать последствия накопления среднеквадратичной ошибки за счет дообучения учительской сети. Однако предложенный подход не позволяет избавиться от проблемы медленного обновления весов при обучении методом обратного распространения ошибки, возникающей из-за наличия операции максимума. Для этого рассмотрим использование непрерывных аппроксимаций операции взятия максимума.

4.3. Непрерывные аппроксимации максимума для БМ сетей

Рассмотрим три различных приближения максимума [15-16]:

$$L_{morph}(x) = \frac{\sum_{i=1}^{n} x_i^{\alpha+1}}{\sum_{i=1}^{n} x^{\alpha}};$$

$$S_{morph}(x) = \frac{\sum_{i=1}^{n} x_i e^{\alpha x_i}}{\sum_{i=1}^{n} e^{\alpha x_i}};$$

$$LSE(x) = \frac{\ln \sum_{i=1}^{n} x_i e^{\alpha x_i}}{\alpha};$$

где x – вектор входных значений, α – температурный параметр, отвечающий за степень приближения к максимуму, n – длина входного вектора. На Рис. 2 представлена средняя абсолютная ошибка отклонения аппроксимации от значения максимума в зависимости от параметра приближения α для усреднённых значений внутри LeNet подобной сети. Можно видеть, что они обеспечивают достаточно точное приближение максимума, а при значении α > 20 – практически совпадают с ним. При использовании этих аппроксимаций при не слишком больших значения параметра α в БМ нейронах на каждом шаге обратного распространения ошибки будет обновляться большее число весов, чем при использовании максимума, что потенциально улучшит обучаемость сети.

4.4. Обучение БМ модели с непрерывной аппроксимацией максимума

Для обучения моделей с помощью описанных методов дистилляции знаний будем использовать непрерывные аппроксимации максимума в два этапа. Первый этап – это обучение модели с аппроксимацией с помощью дистилляции знаний. Поскольку в предложенных аппроксимациях есть операции умножения и деления, то нужен еще один этап, в котором аппроксимация заменяется на точный максимум с сохранением полученных ранее весов. Далее получившаяся

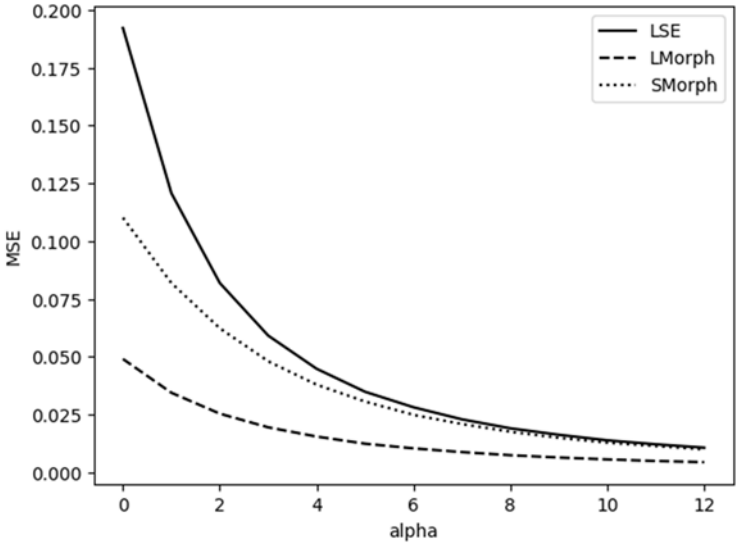


Рис. 2. Среднеквадратичная ошибка (MSE) различных непрерывных аппроксимаций максимума внутри LeNet-подобной сети

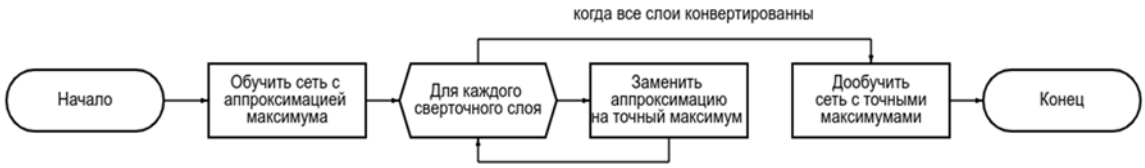


Рис. 3. Алгоритм обучения БМ модели с непрерывной аппроксимацией максимума

сеть дообучается с помощью дистилляции знаний Рис. 3.

5. Результаты экспериментов

Для оценки качества работы представленной модели БМ нейрона, а также предложенного метода обучения были проведены численные эксперименты на нейросетевых моделях с различной архитектурой на выборках MNIST и CIFAR10.

5.1. Результаты обучения с различными непрерывными аппроксимациями максимума

В разделе 4 рассмотрены три непрерывные аппроксимации операции взятия максимума различной вычислительной сложности.

В этом эксперименте выполнено сравнение точности классификации изображений рукописных цифр из выборки MNIST с помощью четырехветочной 5-слойной LeNet-подобной БМ сети, архитектура которой показана на Рис. 4, использующей различные аппроксимации максимума. Для обучения применялись следующие методы:

1. Прямое обучение: классический метод обучения на основе обратного распространения ошибки и градиентных методов.

2. Послойное обучение без фиксации: метод послойного преобразования и дообучения, описанный в разделе 2.

3. Послойное обучение с фиксацией: метод послойного преобразования и дообучения, фиксирующий веса после конвертации и дообучения конкретного слоя.

4. Дистилляция знаний: метод дистилляции знаний, описанный в пункте 4.1.

Параметр аппроксимации определялся следующим образом: производилось обучение сети с обучаемым значением α в каждом слое, после чего значения α были зафиксированы и дальнейшие

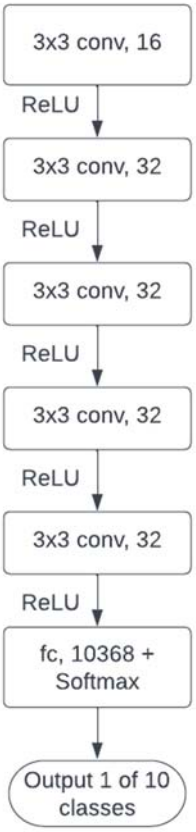


Рис. 4. Использованная нейросетевая архитектура

эксперименты производились с постоянным параметром приближения для каждого слоя.

В Табл. 1 показаны точности классификации для различных аппроксимаций максимума в 4-веточной БМ модели при использовании разных методов обучения. Можно видеть, что LSE аппроксимация показывала наилучшее качество, незначительно превосходящее качество референтной модели. Кроме того, она является самой вычислительно-простой по сравнению с другими аппроксимациями. Поэтому в дальнейших экспериментах использовалась именно она.

Табл. 1. Точность 5-слойной сверточной БМ сети на тестовой выборке MNIST

	Прямое обучение	Послойное обучение без фиксации	Послойное обучение с фиксацией	Дистилляция знаний
Референтная сеть	0.9901	-	-	-
4-БМ	0.1064	0.9648	0.9713	0.9893
4-БМ S-morph	0.9231	0.9666	0.9737	0.9871
4-БМ L-morph	0.1246	0.9361	0.9582	0.9612
4-БМ LSE	**0.9863**	**0.9871**	**0.9922**	**0.9945**

5.2. Обучение 1.5-веточной БМ сети на выборках MNIST и CIFAR10

Так как основной целью создания БМ нейрона является замена операции умножения на операцию сложения, а операцию сложения – на максимум, то в конечной сети не может быть использована непрерывная аппроксимация максимума. В Табл. 2 представлены результаты обучения 1.5-веточной модели с LeNet-подобной архитектурой на выборке MNIST, где **LSE** – модель, использующая LSE, **LSE-MAX** – модель, использующая LSE с последующим преобразованием к точному максимуму (раздел 4.4), **MAX** – модель, использующая только точный максимум в процессе обучения.

Сеть с LSE аппроксимацией показывает отличные результаты с помощью методов обучения на основе дистилляции знаний, в то время как метод прямого обучения демонстрирует низкую точность классификации. При этом методы с точным максимумом с помощью дистилляции знаний не могут добиться тех же значений, что и референтная сеть, но при использовании послойной дистилляции знаний для БМ сети (раздел 4.2) точность становится лучше, чем у референтной сети.

После проведения экспериментов на модели с небольшим числом коэффициентов, была рассмотрена задача классификации изображений выборки CIFAR10 с помощью модели архитектуры ResNet-22 [17] с тремя остаточными блоками, в которых 16, 64 и 256 фильтров соответственно. В каждом блоке 2 сверточных слоя, выход которых суммируется с входом. Результаты этих экспериментов представлены в Табл. 3.

Сеть с LSE аппроксимацией все также показывает лучший результат с помощью обоих методов обучения на основе дистилляции знаний. Методы с точным максимумом показывают схожие результаты, что и на выборке MNIST: дистилляция знаний немного уступает по точности референтной сети, а предложенный метод показывает качество сопоставимое с референтной сетью. При этом использование LSE аппроксимации максимума позволяет упростить сходимость модели и потенциально способно ускорить обучение в сложных задачах по сравнению с использованием послойной дистилляции знаний с точным максимумом.

Табл. 2. Точность 5-слойной сверточной 1.5-веточной БМ сети на тестовой выборке MNIST

	Прямое обучение	Дистилляция знаний	Послойная дистилляция знаний для БМ сетей
Референтная сеть	0.9901	-	-
1.5-БМ LSE	0.1486	**0.9998**	**0.9998**
1.5-БМ LSE-MAX	0.1153	0.9821	**0.9986**
1.5-БМ MAX	0.0926	0.9786	**0.9989**

Табл. 3. Точность ResNet-22 на тестовой выборке CIFAR10

	Прямое обучение	Дистилляция знаний	Послойная дистилляция знаний для БМ сетей
Референтная сеть	0.8643	-	-
1.5-БМ LSE	0.0982	**0.8700**	**0.8700**
1.5-БМ LSE-MAX	0.1033	0.8527	**0.8661**
1.5-БМ MAX	0.1084	0.8514	**0.8669**

Заключение

В этой работе предложена новая 1.5-веточная модель БМ нейрона, повышающая его вычислительную эффективность. Представленная модель позволяет уменьшить количество вычислительных веток нейрона с 4 до 1, и добавляет дополнительный вычислительно-простой шаг в виде подсчета смещения относительно первой четверти. Для этой модели БМ нейрона разработаны методы обучения на основе послойной дистилляции знания и использования непрерывной LSE аппроксимации максимума. Эти методы показали свою эффективность в задачах классификации изображений выборки MNIST для модели архитектуры LeNet: точность классификации с помощью упрощенной БМ модели составила 99.45%, а на выборке CIFAR10 для модели ResNet-22 – точность составила 86.69%. Точность классификации классических сетей составила 99.01% и 86.43% соответственно, что показывает превосходство представленной модели и метода обучения.

Литература

1. Chernyshova Y. S., Sheshkus A. V., Arlazarov V. V. Two-step CNN framework for text line recognition in camera-captured images // IEEE Access. – 2020. – Т. 8. – С. 32587-32600.
2. Kanaeva I. A., Ivanova Y. A., Spitsyn V. G. Deep convolutional generative adversarial network-based synthesis of datasets for road pavement distress segmentation // Computer Optics. – 2021. – Т. 45. – №. 6. – С. 907-916.
3. Das P. A. K., Tomar D. S. Convolutional neural networks based weapon detection: a comparative study // Fourteenth International Conference on Machine Vision (ICMV 2021). – SPIE, 2022. – Т. 12084. – С. 351-359.
4. Bulatov K. et al. Smart IDReader: Document recognition in video stream // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – IEEE, 2017. – Т. 6. – С. 39-44.
5. Zhao Y., Wang D., Wang L. Convolution accelerator designs using fast algorithms // Algorithms. – 2019. – Т. 12. – №. 5. – С. 112.
6. Yao Z. et al. Hawq-v3: Dyadic neural network quantization // International Conference on Machine Learning. – PMLR, 2021. – С. 11875-11886.
7. Tai C. et al. Convolutional neural networks with low-rank regularization // arXiv preprint arXiv:1511.06067. – 2015.
8. Sun X. et al. Pruning filters with L1-norm and standard deviation for CNN compression // Eleventh international conference on machine vision (ICMV 2018). – SPIE, 2019. – Т. 11041. – С. 691-699.
9. You H. et al. Shiftaddnet: A hardware-inspired deep network // Advances in Neural Information Processing Systems. – 2020. – Т. 33. – С. 2771-2783.
10. Chen H. et al. AdderNet: Do we really need multiplications in deep learning? // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2020. – С. 1468-1477.
11. Limonova E. E. et al. Bipolar morphological neural networks: Gate-efficient architecture for computer vision // IEEE Access. – 2021. – Т. 9. – С. 97569-97581.
12. Limonova E. E. Fast and Gate-Efficient Approximated Activations for Bipolar Morphological Neural Networks // Информационные технологии и вычислительные системы. – 2022. – №. 2. – С. 3-10.
13. Hinton G., Vinyals O., Dean J. Distilling the knowledge in a neural network // arXiv preprint arXiv:1503.02531. – 2015.
14. Xu Y. et al. Kernel based progressive distillation for adder neural networks // Advances in Neural Information Processing Systems. – 2020. – Т. 33. – С. 12322-12333.
15. Kirszenberg A. et al. Going beyond p-convolutions to learn grayscale morphological operators // Discrete Geometry and Mathematical Morphology: First International Joint Conference, DGMM 2021, Uppsala, Sweden, May 24–27, 2021, Proceedings. – Cham : Springer International Publishing, 2021. – С. 470-482.
16. Calafiore G. C., Gaubert S., Possieri C. A universal approximation result for difference of log-sum-exp neural networks // IEEE transactions on neural networks and learning systems. – 2020. – Т. 31. – №. 12. – С. 5603-5612.
17. He K. et al. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.

Зингеренко Михаил Владимирович. Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Московская область, Россия. Студент аспирант. ООО «Смарт Энджинс Сервис», Москва, Россия. Лаборант-программист. Область научных интересов: нейронные сети, обработка изображений. E-mail: m.zingerenko@smartengines.com

Лимонова Елена Евгеньевна. Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия. Научный сотрудник, кандидат технических наук. ООО «Смарт Энджинс Сервис» Лаборант-программист. Область научных интересов: нейронные сети, обработка изображений, распознавания образов на мобильных устройствах. E-mail: limonova@smartengines.com

Layer-Wise Knowledge Distillation for Simplified Bipolar Morphological Neural Networks

M. V. Zingerenko [I, II], E. E. Limonova [II, III]

[I] Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Moscow Region, Russia
[II] Smart Engines Service LLC, Moscow, Russia
[III] Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

Abstract. Various neuron approximations can be used to reduce the computational complexity of neural networks. One such approximation based on summation and maximum operations is a bipolar morphological neuron. This paper presents an improved structure of the bipolar morphological neuron that enhances its computational efficiency and a new approach to training based on continuous approximations of the maximum and knowledge distillation. Experiments were conducted on the MNIST dataset using a LeNet-like neural network architecture and on the CIFAR10 dataset using a ResNet-22 model architecture. The proposed training method achieves 99.45% classification accuracy on the LeNet-like model, with the same accuracy of the classical network, and 86.69% accuracy on the ResNet-22 model, compared to 86.43% accuracy of the classical model. The results show that the proposed method with log-sum-exp (LSE) approximation of the maximum and layer-by-layer knowledge distillation, allows for a simplified bipolar morphological network that is not inferior to classical networks.
Keywords: bipolar morphological networks, approximations, artificial neural networks, computational efficiency.

DOI 10.14357/20718632230305

References

1. Chernyshova YS, Sheshkus AV, Arlazarov VV. Two-step CNN framework for text line recognition in camera-captured images. IEEE Access. 2020;8:32587-600.
2. Kanaeva I, Ivanova YA, Spitsyn V. Deep convolutional generative adversarial network-based synthesis of datasets for road pavement distress segmentation. Computer Optics. 2021;45(6):907-16.
3. Das PAK, Tomar DS. Convolutional neural networks based weapon detection: a comparative study. In: Fourteenth International Conference on Machine Vision (ICMV 2021). vol. 12084. SPIE; 2022. p. 351-9.
4. Bulatov K, Arlazarov VV, Chernov T, Slavin O, Nikolaev D. Smart IDReader: Document recognition in video stream. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). vol. 6. IEEE; 2017. p. 39-44.
5. Zhao Y, Wang D, Wang L. Convolution accelerator designs using fast algorithms. Algorithms. 2019;12(5):112.
6. Yao Z, Dong Z, Zheng Z, Gholami A, Yu J, Tan E, et al. Hawq-v3: Dyadic neural network quantization. In: International Conference on Machine Learning. PMLR; 2021. p. 11875-86.
7. Tai C, Xiao T, Zhang Y, Wang X, et al. Convolutional neural networks with low-rank regularization. arXiv preprint arXiv:151106067. 2015.
8. Sun X, Zhou D, Pan X, Zhong Z, Wang F. Pruning filters with L1-norm and standard deviation for CNN compression. In: Eleventh international conference on machine vision (ICMV 2018). vol. 11041. SPIE; 2019. p. 691-9.
9. You H, Chen X, Zhang Y, Li C, Li S, Liu Z, et al. Shiftaddnet: A hardware-inspired deep network.

Advances in Neural Information Processing Systems. 2020;33:2771-83.
10. Chen H, Wang Y, Xu C, Shi B, Xu C, Tian Q, et al. AdderNet: Do we really need multiplications in deep learning? In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 1468-77.
11. Limonova EE, Alfonso DM, Nikolaev DP, Arlazarov VV. Bipolar morphological neural networks: Gate-efficient architecture for computer vision. IEEE Access. 2021;9:97569-81.
12. Limonova EE. Fast and Gate-Efficient Approximated Activations for Bipolar Morphological Neural Networks. Informacionnye tekhnologii I vichslitel'nye sistemy 2022;(2):3-10.
13. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv preprint arXiv:150302531. 2015.
14. Xu Y, Xu C, Chen X, Zhang W, Xu C, Wang Y. Kernel based progressive distillation for adder neural networks. Advances in Neural Information Processing Systems. 2020;33:12322-33.
15. Kirszenberg A, Tochon G, Puybareau E´, Angulo J.Going beyond p-convolutions to learn grayscale morphological operators. In: International Conference on Discrete Geometry and Mathematical Morphology. Springer; 2021. p. 470-82.
16. Calafiore GC, Gaubert S, Possieri C. A universal approximation result for difference of log-sum-exp neural networks. IEEE transactions on neural networks and learning systems. 2020;31(12):5603-12.
17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770-8.

Zingerenko Mikhail V. Moscow Institute of Physics and Technology (National Research University), Dolgoprudny, Moscow Region, Russia. PhD Student; Smart Engines Service LLC, Moscow, Russia. Programmer. Research interests: neural networks, image processing, pattern recognition on mobile devices. E-mail: m.zingerenko@smartengines.com

Limonova Elena E. Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia, PhD, researcher; LLC "Smart Engines Service", Programmer. Research interests: neural networks, image processing, pattern recognition on mobile devices. E-mail: limonova@smartengines.com