

Автоматическая настройка операции пулинга в сверточных нейронных сетях для классификации угольных пород

О.А. Козлова¹, В.В. Китов^{1,2}✉

¹ Московский государственный университет имени М.В. Ломоносова, г. Москва, Российская Федерация

² Российский экономический университет имени Г.В. Плеханова, г. Москва, Российская Федерация

✉ v.v.kitov@yandex.ru

Резюме: Сверточные нейронные сети автоматизируют задачи обработки изображений, такие как классификация, сегментация, обнаружение объектов, перенос стиля и т. д. Эти сети активно применяются в угольной промышленности для автоматической классификации угольных пород с высокой точностью на основе необработанных изображений. Точная классификация угольных пород важна для оценки качества угля, оптимизации процессов его добычи, обогащения и переработки. Основными математическими операциями сверточных сетей являются свертка и пулинг. В статье рассматривается обобщение операции пулинга. Обычно тип пулинга задается заранее некоторой агрегирующей операцией – усреднением (average pooling) или взятием максимума (max pooling). Размер агрегируемой области также задается заранее. Тип пулинга и размер агрегируемой области существенно влияют на качество обработки изображений угольных пород. В статье предложено несколько параметрических обобщений операции пулинга, которые охватывают average и max pooling как частные случаи. Также для max pooling предложено параметрическое обобщение, позволяющее варьировать размер области агрегации. Параметры предложенных обобщений пулинга обучаются автоматически вместе с остальными весами сети.

Ключевые слова: обобщение пулинга, автонастраиваемый пулинг, поиск нейронной архитектуры, классификация изображений, угольные породы, угольная промышленность

Благодарности: Работа выполнена в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования РФ, проект «Модели, методы и алгоритмы искусственного интеллекта в задачах экономики для анализа и переноса стиля многомерных наборов данных, прогнозирования временных рядов и построения рекомендательных систем», грант № FSSW-2023-0004.

Для цитирования: Козлова О.А., Китов В.В. Автоматическая настройка операции пулинга в сверточных нейронных сетях для классификации угольных пород. *Горная промышленность*. 2024;(6):125–134. <https://doi.org/10.30686/1609-9192-2024-6-125-134>

Automatic tuning of the pooling operation in convolutional neural networks for coal rock classification

О.А. Kozlova¹, V.V. Kitov^{1,2}✉

¹ Lomonosov Moscow State University, Moscow, Russian Federation

² Plekhanov Russian University of Economics, Moscow, Russian Federation

✉ v.v.kitov@yandex.ru

Keywords: Convolutional neural networks are used to automate image processing tasks such as classification, segmentation, object detection, style transfer, etc. These networks are actively applied in coal mining industry for automatic classification of coal rocks with high accuracy based on raw images. Accurate classification of the coal rocks is important for coal quality assessment, optimization of coal mining, preparation and processing. The main mathematical operations of convolutional networks are convolution and pooling. The paper discusses a generalization of the pooling operation. Usually the type of pooling is specified in advance by some aggregation operation, i.e. the average pooling or max pooling. The size of the aggregated area is also specified in advance. The type of pooling and the size of the aggregated region significantly affect the quality of coal-bed image processing. This paper proposes several parametric generalizations of the pooling operation, which cover the average and max pooling as special cases. A parametric generalization is also proposed for max pooling, which allows to vary the size of the aggregation region. Parameters of the proposed pooling generalizations are automatically trained along with the rest of the network weights.

Keywords: pooling generalization, auto-tuned pooling, neural architecture search, image classification, coal rocks, coal industry

Acknowledgments: The work was carried out within the framework of the state assignment in the field of scientific activity of the Ministry of Science and Higher Education of the Russian Federation, Project entitled “Models, methods and algorithms of artificial intelligence in economic problems for analyzing and transferring the style of multidimensional data sets, forecasting time series and building recommendation systems”, Grant No. FSSW-2023-0004.

For citation: Kozlova O.A., Kitov V.V. Automatic tuning of the pooling operation in convolutional neural networks for coal rock classification. *Russian Mining Industry*. 2024;(6):125–134. (In Russ.) <https://doi.org/10.30686/1609-9192-2024-6-125-134>

Введение

Большинство задач обработки изображений, таких как классификация изображений, семантическая сегментация, обнаружение объектов, перенос стиля, повышение разрешения изображений и т.д., основаны на сверточных нейронных сетях. В угольной промышленности эти сети используются для автоматической классификации угольных пород по изображениям, что является важной задачей для оценки качества угля, планирования горных работ, управления процессами обогащения и переработки угля.

Существует несколько наборов данных с изображениями угольных пород, такие как Coal Rock Images¹, содержащий 10 классов угольных пород из различных угольных шахт Китая, Coal and Rock Images² с 12 классами угольных и вмещающих пород, и CoalRockImage5k³, включающий 5000 изображений угольных пород из российских месторождений, разделенных на 5 классов по содержанию золы. Последний набор данных был исследован в работе [1], где с помощью сверточных нейронных сетей была достигнута точность классификации угольных пород по изображениям 95%. Применение нейросетевых методов позволило значительно повысить точность и автоматизировать процесс оценки качества угля по сравнению с традиционными методами, основанными на ручном визуальном анализе ядерного материала.

Основными операциями сверточных нейронных сетей являются свертки, нелинейности и операции пулинга. Полносвязные слои (многослойный перцептрон) или другие методы машинного обучения могут впоследствии применяться в конце сверточной нейронной сети. Популярность сверточных нейронных сетей объясняется тем, что они способны автоматически извлекать сложные и нетривиальные признаки из изображений угольных пород, которые лучше всего решают конечную задачу классификации, без переобучения, поскольку операция свертки содержит мало параметров (из-за общности весов и локального характера действия свертки) и операции нелинейности, такие как ReLU, тогда как стандартные операции пулинга вообще не содержат обучаемых параметров.

Однако операция пулинга содержит гиперпараметры (устанавливаемые пользователем) – тип используемой операции агрегации (усреднение при усредняющем пулинге и взятие максимума при максимизирующем пулинге), а также размер области, по которой выполняется агрегация (2x2, 4x4 и т.д.). Эти конфигурации существенно влия-

ют на качество извлекаемых признаков из изображений угольных пород и на точность решения задачи их классификации. Поэтому эти гиперпараметры обычно настраиваются вручную методом проб и ошибок, что замедляет процесс построения модели для анализа угольных пород.

Материалы и методы

Классические типы пулинга

Максимизирующий пулинг (max-pooling) и усредняющий пулинг (average pooling) являются двумя наиболее распространенными методами агрегации в сверточных нейронных сетях, применяемых в том числе для анализа угольных пород. Максимизирующий пулинг, предложенный в [2], извлекает наибольшее значение из нескольких пикселей в определенной области изображения угольной породы и использует его для создания новой уменьшенной карты признаков. В свою очередь, усредняющий пулинг [3] вычисляет среднее значение всех пикселей в области и использует его для создания новой карты признаков. Оба метода помогают снизить вычислительную нагрузку сети за счет уменьшения размерности данных, что может ускорить обучение на большом объеме изображений угольных пород и уменьшить вероятность переобучения. Однако выбор между этими методами зависит от конкретной задачи и типа угольных пород. Ниже приведены формулы для максимизирующего пулинга и усредняющего пулинга соответственно, применяемых к промежуточному представлению изображения угольной породы $x \in R^{C \times H \times W}$ (C – число каналов признаков, а H, W – высота и ширина соответственно):

$$y_{cij} = \max_{(p,q) \in R_{ij}} (x_{cpq}); \quad y_{cij} = \frac{1}{|R_{ij}|} \sum_{(p,q) \in R_{ij}} x_{cpq}, \quad (1)$$

где y_{cij} – выход оператора пулинга на канале признаков c и пространственной позиции i, j ; x_{cpq} – входной элемент канала c в позиции p, q внутри области пулинга R_{ij} , которая представляет собой локальную окрестность вокруг позиции i, j на изображении угольной породы, а $|R_{ij}|$ – размер (число элементов) соответствующей области. Примеры применения max и average пулинга к изображению угольной породы показаны на рис. 1.

Стандартный пулинг требует ручного выбора его типа – усреднение или максимизация. Но даже выбор типа пулинга не гарантирует оптимальности для задачи классификации угольных пород, поскольку оптимальная операция агрегации может быть не точно усреднением или максимизацией, а некоторой промежуточной операцией между ними. Поэтому представляют интерес промежуточные и автоматически настраиваемые типы пулинга, представляющие

1 Dataset – Minerals identification Available at: <https://www.kaggle.com/datasets/asiadubrempong/minerals-identification-dataset> (assessed: 01.09.2024).

2 Dataset – Minerals classifications Available at: <https://www.kaggle.com/datasets/youcefattallah97/minerals-identification-classification> (assessed: 01.09.2024).

3 Dataset – All minerals pack Available at: <https://www.kaggle.com/datasets/sergeynesteruk/minerals> (assessed: 01.09.2024).

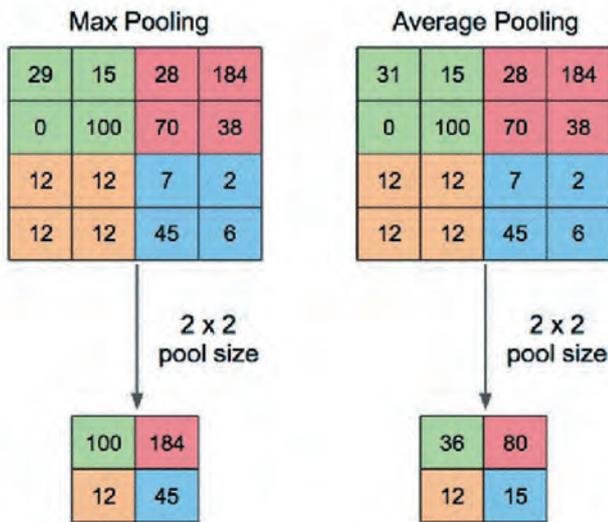


Рис. 1
Примеры усредняющего и максимизирующего пулинга на изображении угольной породы

Fig. 1
Examples of the average and max pooling on a coal rock image

max-pooling и усредняющий пулинг как частные случаи и позволяющие адаптировать операцию пулинга под конкретный тип угольных пород.

Известен существующий гибридный подход [4–6], называемый смешанным пулингом (mixed pooling), сочетающий усредняющий пулинг и максимизирующий пулинг. Он вычисляется по формуле:

$$y_{cij} = \lambda \cdot \max(x_{cij}) + (1 - \lambda) \cdot \text{avg}(x_{cij}), \quad (2)$$

Параметр λ в (2) принимает значения от 0 до 1 и отвечает за долю использования максимизации и усреднения при пулинге. В некоторых работах [5] предлагается выбирать значение этого параметра случайно из интервала [0, 1], но такой подход работает менее точно по сравнению с настройкой параметра λ методом обратного распространения ошибки вместе с остальными весами сети. Смешанный пулинг применялся в [6] для классификации угольной породы по 10 классам из набора данных Coal Rock Images⁴ и показал более высокую точность по сравнению со стандартным максимизирующим и усредняющим пулингом.

Для обучения коэффициента смешивания операций максимизации и усреднения можно выбирать разные стратегии, например, в [7] вводились два настраиваемых параметра λ_1 и λ_2 , $\lambda = \frac{\lambda_1^2}{\lambda_1^2 + \lambda_2^2}$. А в [4] смешивание настраивалось с помощью одного параметра λ_1 и сигмоидной функции от него:

$$\lambda = \sigma(\lambda_1) = \frac{1}{1 + \exp(-\lambda_1)}$$

Другим вариантом автонастройки типа пулинга является L_p пулинг [8; 9]. Он основан на суммировании элементов по определенной области, затем возведении в степень p и делении на размер этой области:

$$y_{cij} = \left(\frac{1}{|R_{ij}| \sum_{(s,q) \in R_{ij}} x_{csq}^p} \right)^{\frac{1}{p}}. \quad (3)$$

В этом случае p является гиперпараметром, определяющим степень важности больших значений, так как при $p \rightarrow \infty$ (3) сходится к max-пулингу, а при $p = 1$ операция (3) действует как усредняющий пулинг. Для корректности операции перед использованием этого пулинга применяется нелинейная активация ReLU [10] или другая, после чего значения x_{csq} становятся неотрицательными. L_p пулинг был использован в [9] для классификации угольных пород из набора Coal and Rock Images⁵ и показал лучшие результаты при $p = 3$ по сравнению с усредняющим и максимизирующим пулингом.

Существует также подход, показывающий улучшение качества на задаче классификации угольных пород, если брать пулинг как конкатенацию вдоль каналов максимизирующего и усредняющего пулинга [11]. Однако недостатком такого подхода является значительное увеличение числа параметров нейронной сети и объема вычислений. Еще одним подходом к обучению параметров пулинга является LEAP пулинг [12], который работает очень похоже на свёрточный слой в нейронной сети. Единственное отличие в том, что при замене пулинга сверткой не используется информация из других каналов:

$$y_{cij} = \sum_{(s,q) \in R_{ij}} w_{csq} x_{csq} \quad (4)$$

В табл. 1 приведено сравнение точности классификации угольных пород различными существующими методами пулинга на трех наборах данных. Смешанный пулинг и L_p пулинг во всех случаях превосходят по точности стандартные максимизирующий и усредняющий пулинг, подтверждая перспективность автонастраиваемых типов пулинга для анализа угольных пород.

Таблица 1
Сравнение точности классификации угольных пород различными методами пулинга, %

Table 1
Comparison of the coal rock classification accuracy of different pooling methods, %

Метод пулинга	Coal Rock Images	Coal and Rock Images	CoalRockImage5k
Max pooling	91,5	88,3	94,1
Avg pooling	92,1	89,5	94,6
Mixed pooling	93,4	90,2	95,3
L_p pooling	93,8	90,7	95,5
LEAP pooling	92,4	89,9	94,9

Предложенные методы

1. Softmax pooling. Для обобщения усредняющего и максимизирующего пулинга предлагается использовать локальное взвешенное усреднение, где в качестве весов внутри области выступает функция softmax от активаций с автонастраиваемой температурой. Мы называем такой пулинг softmax pooling.

4 Dataset – Minerals identification Available at: <https://www.kaggle.com/datasets/asiedubrempong/minerals-identification-dataset> (assessed: 01.09.2024).

5 Dataset – Minerals classifications Available at: <https://www.kaggle.com/datasets/youcefattallah97/minerals-identification-classification> (assessed: 01.09.2024).

$$y_{cij} = \sum_{(p,q) \in R_{ij}} \frac{x_{cpq} \left(e^{-\frac{x_{cpq}}{T}} \right)}{\sum_{(p',q') \in R_{ij}} e^{-\frac{x_{cp'q'}}{T}}} \quad (5)$$

где $T \geq 0$ – так называемый параметр «температуры».

Можно заметить, что при $T \rightarrow \infty$ значение коэффициентов при x_{cpq} стремится к равномерным весам $\frac{1}{|R_{ij}|}$, что соответствует усредняющему пулингу. При $T \rightarrow 0$ мы получаем максимизирующий пулинг как частный случай. Действительно, делая замену $T_1 = \frac{1}{T}$ и обозначая $x_{max} = \max_{(p,q) \in R_{ij}} (x_{cpq})$, получаем для каждого канала c :

$$w_{pq} = \lim_{T \rightarrow 0} \frac{\exp\left(\frac{x_{cpq}}{T}\right)}{\sum_{(p',q') \in R_{ij}} \exp\left(\frac{x_{cp'q'}}{T}\right)} =$$

$$\lim_{T' \rightarrow \infty} \frac{\exp(x_{cpq} \cdot T')}{\sum_{(p',q') \in R_{ij}} \exp(x_{cp'q'} \cdot T')} =$$

$$= \exp((x_{cpq} - x_{max}) \cdot T') \frac{\lim_{T' \rightarrow \infty}}{\sum_{(p',q') \in R_{ij}} \left(\exp((x_{cp'q'} - x_{max}) \cdot T') \right)}$$

$$= \begin{cases} 1, & x_{cpq} = x_{max} \\ 0, & otherwise \end{cases}$$

Таким образом, softmax pooling представляет собой гибкое обобщение усредняющего и максимизирующего пулинга, автоматически настраивающееся под особенности классифицируемых угольных пород.

2. Геометрический и гармонический пулинг. Также предлагается обобщить максимизирующий и усредняющий пулинги путем взятия среднего геометрического (GeometricPool) и среднего гармонического (HarmonicPool) над ними, вычисляемых по формулам (6) и (7) соответственно:

$$y_{kij} = \max(x_{cij})^\lambda \cdot \text{avg}(x_{cij})^{1-\lambda} \quad (6)$$

$$y_{cij} = \frac{1}{\frac{\lambda}{\max(x_{cij})} + \frac{1-\lambda}{\text{avg}(x_{cij})}} \quad (7)$$

Параметр $\lambda \geq 0$ предлагается настраивать вместе с остальными параметрами нейронной сети методом обратного распространения ошибки.

Геометрический пулинг можно интерпретировать как взвешенное произведение максимизирующего и усредняющего пулингов, настраиваемое в зависимости от типа угольной породы. Гармонический пулинг представляет собой взвешенное гармоническое среднее, которое больше штрафует малые значения пулингов по сравнению со средним арифметическим и также адаптируется под конкретные изображения угольных пород.

3. Автоматический выбор размера области пулинга. В данном разделе предлагается метод максимизирующего пулинга с автоматическим выбором размера области, по которой применяется максимизация. Задаются две

области R_1 и R_2 , причем $R_1 \subset R_2$, например, области 2×2 и 4×4 . Вместо заранее заданной одной конкретной области предложенный метод автоматически выбирает одну из заданных областей или некоторый промежуточный вариант в зависимости от настраиваемого параметра λ , как показано в формуле:

$$y_{cij} = \begin{cases} \lambda \cdot \max_{(p,q) \in R_{ij}^1} x_{cpq} + (1-\lambda) \max_{(p,q) \in R_{ij}^2} x_{cpq} & \text{if } \max_{(p,q) \in R_{ij}^2} x_{cpq} \notin R_{ij}^1 \\ \lambda \cdot \max_{(p,q) \in R_{ij}^1} x_{cpq} & , \text{ otherwise} \end{cases} \quad (8)$$

Параметр λ автоматически настраивается в ходе обучения нейросети таким образом, чтобы выбрать оптимальный размер области пулинга для классификации конкретных угольных пород.

Заметим, что важно явно разделять на два случая, чтобы не было случайного изменения параметра λ в случае, когда максимальное значение внутри областей лежит в меньшей из областей. Иначе любое значение параметра λ будет подходить, и параметр может неверно подстраиваться в ходе оптимизации.

Результаты

В данном разделе различные методы пулинга сравниваются на следующих задачах классификации изображений угольных пород:

Coal Rock Images ⁶: изображения размером 200x200 пикселей с 3 цветовыми каналами (RGB). Общее число изображений в выборке составляет 10 000, из которых 8000 отведено под обучение, а оставшиеся 2000 – под тестовую выборку. Каждый из 10 классов угольных пород (антрацит, бурый уголь, углистый сланец, алевролит и др.) содержит по 1000 изображений в наборе данных.

Coal and Rock Images ⁷: 10 000 цветных изображений размером 224x224 пикселя, разделенных на 12 классов угольных и вмещающих пород, представленных 800–900 изображениями в каждом классе. Все изображения разделены на 8000 изображений для обучения и 2000 изображений для тестирования.

CoalRockImage5k ⁸: набор данных, содержащий 5000 изображений угольных пород размером 300x300 пикселей. Присутствует 5 классов в зависимости от содержания золы (менее 10%, 10–20%, 20–30%, 30–40%, более 40%), каждый из которых представлен 1000 изображениями. Выборка разделена на 4000 изображений для обучения и 1000 изображений для тестирования.

Во всех экспериментах использовалась сверточная нейронная сеть архитектуры VGG-11 [1], показавшая хорошие результаты на различных задачах классификации изображений. Модели реализовывались с помощью Python библиотеки PyTorch и оптимизировались методом Adam со скоростью обучения $lr = 10^{-5}$ и функцией потерь кросс-энтропия. Изображения масштабировались и обрезались до одинакового разрешения 224x224 пикселя для возможности использования одной и той же модели независимо от набора данных. Структура модели показана на рис. 2.

Модель обучалась с различными типами пулинга. Обучение происходило до достижения значением точности

⁶ Dataset – Minerals identification Available at: <https://www.kaggle.com/datasets/asiedubrempong/minerals-identification-dataset> (assessed: 01.09.2024).

⁷ Dataset – Minerals classifications Available at: <https://www.kaggle.com/datasets/youcefattallah97/minerals-identification-classification> (assessed: 01.09.2024).

⁸ Dataset – All minerals pack Available at: <https://www.kaggle.com/datasets/sergynesteruk/minerals> (assessed: 01.09.2024).

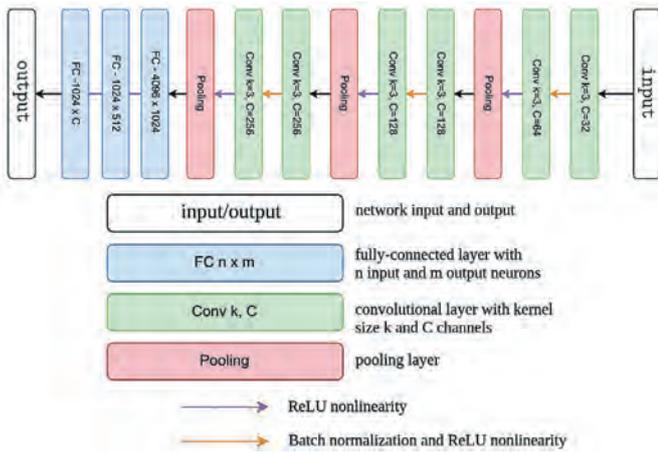


Рис. 2
Структура модели классификации угольных пород

Fig. 2
Structure of the coal rock classification model

стабильного уровня, причем число итераций обучения внутри каждого набора данных было одинаковым для разных типов пулинга для сопоставимости результатов.

Сравнение различных типов пулинга

Ниже приведены сокращения для типов пулинга, которые сравнивались друг с другом в модели на рис. 2. Размер области пулинга везде задавался равным 2x2.

Значения точности на тесте для различных типов пулинга представлены в табл. 2.

Таблица 2
Точность классификации угольных пород для различных типов пулинга, %

Table 2
Accuracy of coal rock classification for different pooling types, %

Метод пулинга	Coal Rock Images	Coal and Rock Images	CoalRockImage5k
original-max	94,8	90,6	95,9
original-avg	95,3	91,5	96,2
concat	96,1	92,3	96,8
mix-pool 1	96,7	93,2	97,4
mix-pool 2	96,5	92,9	97,1
softmax pool	97,2	93,6	97,8
geometric mean	97,5	94,1	98,2
harmonic mean	97,3	93,9	98,0

Как видно из таблицы, смешивание усредняющего и максимизирующего пулинга приводит к устойчивому повышению точности классификации угольных пород по сравнению с использованием усредняющего и максимизирующего пулинга по отдельности. Более того, предложенные методы улучшают качество по сравнению с существующими обобщениями пулинга. Наилучшим оказался геометрический пулинг, что можно объяснить тем, что он добавляет самое сильное нелинейное преобразование к average и max пулингам, как показано на рис. 3. Это позволяет лучше учитывать характерные особенности текстуры и структуры угольных пород на изображениях.

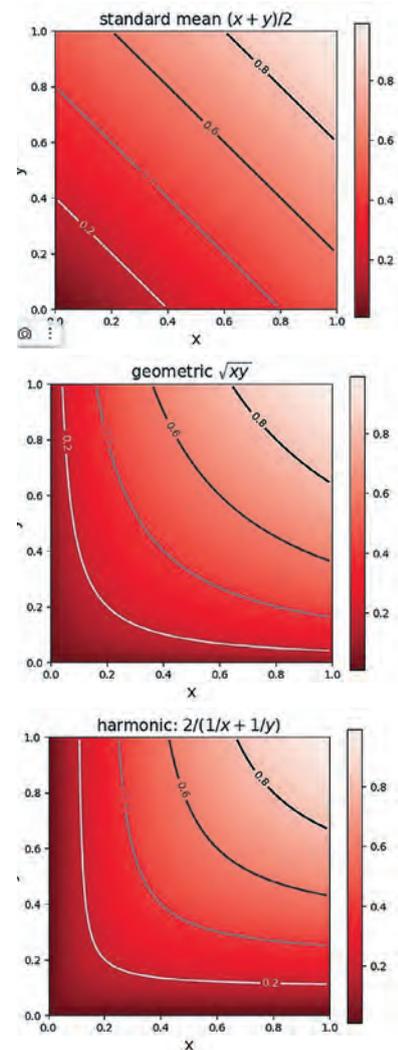


Рис. 3
Линии уровня для среднего арифметического, среднего геометрического и среднего гармонического с равномерными весами. Линии уровня среднего гармонического наиболее выпуклые, что добавляет преобразованию наибольшую нелинейность

Fig. 3
Level lines for arithmetic mean, geometric mean and harmonic mean with equal weights. The level lines for the harmonic mean are the most convex, which adds the most nonlinearity to the transformation

Автоматическое определение размера пулинга

Ниже приведены сокращения для типов пулинга во второй серии экспериментов, проверяющих качество работы предложенного подхода автоматического определения размера области максимизирующего пулинга:

- original-max – стандартный максимизирующий пулинг 2x2;
- original-max 4x4 – стандартный максимизирующий пулинг 4x4;
- size adaptive – предложенный пулинг (формула (8)), использующий автонастраиваемую область между областью размера 2 x 2 и областью размера 4 x 4.

Для обработки ситуаций, когда ядро пулинга выходит за границу входного тензора во втором и третьем методах, использовалось повторяющееся дополнение (repeated padding – повторение значения ближайшего корректного пикселя). Шаг пулинга (смещение области пулинга) для всех методов задавался равным 2 пикселям.

Таблица 3
Точность классификации угольных пород для максимизирующих пулингов разного размера, %

Метод пулинга	Coal Rock Images	Coal and Rock Images	CoalRockImage5k
original-max	94,8	90,6	95,9
original-max 4x4	95,2	91,1	96,3
size adaptive	96,3	92,4	97,1

Table 3
Accuracy of the coal rock classification for max pullings of different sizes, %

Таблица 4
Сравнение времени построения прогноза в миллисекундах

Метод пулинга	Coal Rock Images	Coal and Rock Images	CoalRockImage5k
original-max	12,3	14,1	18,5
original-avg	12,5	14,3	18,7
concat	19,4	22,3	29,1
mix-pool 1	13,2	15,1	19,8
mix-pool 2	13,1	15,0	19,6
softmax pool	14,5	16,7	21,9
geometric mean	13,8	15,9	20,7
harmonic mean	13,6	15,6	20,4
size adaptive	13,5	15,5	20,3

Table 4
Comparison of the prediction time in milliseconds

Результаты сравнительных экспериментов представлены в табл. 3. Предложенная модель пулинга с автонастраиваемым размером показывает наилучшее качество для всех наборов данных угольных пород. Это можно объяснить тем, что оптимальный размер области максимизации зависит от масштаба и детальности характерных структур на изображениях разных типов угольных пород. Автоматическая настройка позволяет адаптировать размер пулинга индивидуально для каждого набора данных.

На рис. 4 показаны графики изменения настраиваемых параметров λ относительно числа эпох на наборе данных Coal Rock Images. Номера пулингов соответствуют пулингам модели, показанной на рис. 2. Первый пулинг расположен ближе всего ко входному изображению, второй – посередине, а третий находится в конце архитектуры.

Видно, что базовые методы усредняющего и максимизирующего пулинга ожидаемо работают быстрее остальных, так как они самые простые. Долше всего работает пулинг-конкатенация, поскольку он удваивает число выходов пулинга, а остальные обобщения базовых пулингов не приводят к столь существенному замедлению. При этом затраты вычислительного времени на новые методы пулинга достаточно малы по сравнению с получаемым повышением точности классификации угольных пород.

Эксперименты на различных марках угля

Для более детального анализа эффективности предложенных методов пулинга были проведены эксперименты на наборах данных с изображениями различных марок угля. Использовались следующие общепринятые марки угля:

- Антрацит (А) – уголь с наивысшей степенью углефикации, характеризующийся высокой теплотворной способностью, низким содержанием влаги и летучих веществ.
- Коксующийся уголь (К) – уголь, пригодный для производства металлургического кокса. Характеризуется низким содержанием золы и серы, высоким содержанием углерода и летучих веществ.
- Жирный уголь (Ж) – уголь с высоким выходом летучих веществ, используется для производства кокса и энергетических целей.

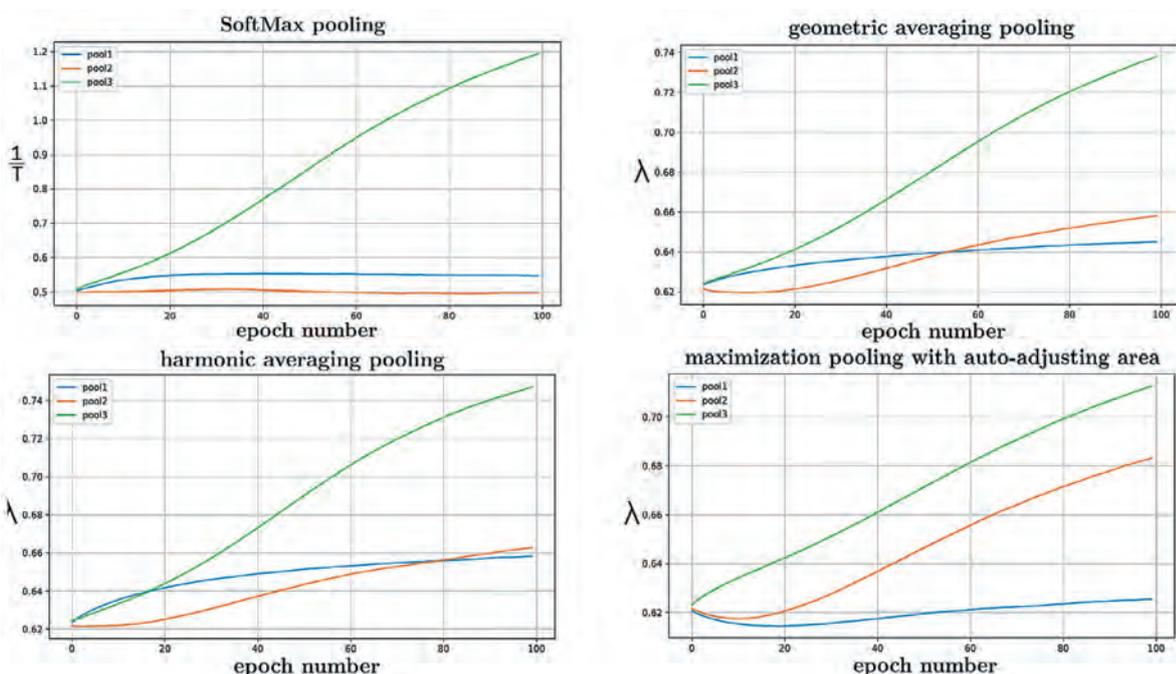


Рис. 4
Динамика настраиваемых параметров на наборе Coal Rock Images

Fig. 4
Changes in the tunable parameters on the coal rock images set

Таблица 5
Точность классификации углей по маркам для различных типов пулинга, %

Метод пулинга	Антрацит (А)	Коксующийся (К)	Жирный (Ж)	Газовый (Г)	Длиннопламенный (Д)	Бурый (Б)
original-max	98,2	96,7	95,9	94,5	92,1	89,8%
original-avg	98,5	97,1	96,3	94,9	92,6	90,4%
softmax pool	99,1	97,9	97,4	96,2	94,5	92,7%
geometric mean	99,3	98,2	97,7	96,6	94,9	93,2%
harmonic mean	99,2	98,1	97,5	96,4	94,7%	93,0%
size adaptive	99,0	97,8	97,2	96,0	94,2	92,4%

Table 5
Accuracy of the coal classification by grades for different types of pooling, %

- Газовый уголь (Г) – уголь с высоким содержанием летучих веществ и низким содержанием углерода, используется для энергетических целей и производства синтетического газа.
- Длиннопламенный уголь (Д) – уголь с высоким выходом летучих веществ и низкой теплотворной способностью, в основном используется как энергетическое и бытовое топливо.
- Бурый уголь (Б) – уголь низкой степени углефикации с высоким содержанием влаги и низкой теплотворной способностью. Используется для энергетических целей и производства химических продуктов.

Для каждой марки угля был собран набор данных, состоящий из 10 000 RGB изображений кусков угля размером 227x227 пикселей, сфотографированных в лабораторных условиях. Наборы данных были разделены на обучающую (8000 изображений) и тестовую (2000 изображений) выборки.

Модель на основе архитектуры VGG-11 обучалась на каждом наборе данных отдельно в течение 50 эпох с использованием различных типов пулингов. Результаты классификации углей по маркам для предложенных методов пулинга в сравнении с базовыми представлены в табл. 5.

Из табл. 5 видно, что предложенные методы пулинга превосходят по точности базовые для всех рассмотренных марок угля. При этом наилучшие результаты показали геометрический и гармонический пулинг. Стоит отметить, что точность классификации снижается с уменьшением

степени метаморфизма угля – от антрацита к бурому углю. Для анализа зависимости точности классификации от количества обучающих изображений были проведены эксперименты с постепенным увеличением размера обучающей выборки от 1000 до 8000 изображений. Точность классификации возрастает с увеличением обучающей выборки для всех методов пулинга. При этом предложенные методы позволяют достигать высокой точности при меньшем числе обучающих изображений по сравнению с базовыми методами. Так, для достижения точности 95% на антраците геометрическому пулингу требуется около 3000 обучающих изображений, в то время как стандартному максимизирующему пулингу – около 4500.

Для изучения влияния качества изображений на точность классификации была проведена серия экспериментов с искусственным ухудшением исходных изображений углей. Ухудшение производилось путем последовательного применения следующих искажений:

- Гауссов шум с различным стандартным отклонением (от 0 до 50);
- Размытие по Гауссу с различным радиусом (от 0 до 5);
- Изменение контрастности (умножение на коэффициент от 0.5 до 1.5);
- Изменение яркости (прибавление константы от -50 до 50).

Для каждого типа и уровня искажений производилось обучение и тестирование модели классификации. Результаты для гауссова шума и размытия по Гауссу приведены в табл. 6 и 7. Для краткости указаны только результаты для standard-max, softmax pool и geometric mean пулингов.

Таблица 6
Точность классификации углей при наличии гауссова шума, %

Марка угля	Пулинг	Стандартное отклонение шума					
		0	10	20	30	40	50
Антрацит	std-max	98,2	97,1	95,4	92,3	87,6	81,2
	softmax	99,1	98,6	97,8	96,4	94,0	90,3
	geo-mean	99,3	98,9	98,2	97,1	95,2	92,1
Коксующийся	std-max	96,7	95,2	92,9	89,1	83,7	76,8
	softmax	97,9	97,2	96,0	94,1	91,2	87,0
	geo-mean	98,2	97,7	96,7	95,2	92,8	89,3
Бурый	std-max	89,8	87,2	83,0	77,1	69,4	60,2
	softmax	92,7	91,4	89,3	86,1	81,5	75,3
	geo-mean	93,2	92,2	90,5	87,9	84,0	78,6

Table 6
Accuracy of the coal classification in the presence of the Gaussian noise, %

Таблица 7
Точность классификации углей при наличии размытия по Гауссу, %

Table 7
Accuracy of the coal classification in the presence of the Gaussian blur, %

Марка угля	Пулинг	Радиус размытия					
		0	1	2	3	4	5
Антрацит	std-max	98,2	97,6	96,0	94,0	90,2	85,1
	softmax	99,1	98,8	98,1	96,9	94,8	91,7
	geo-mean	99,3	99,1	98,6	97,6	95,9	93,3
Коксующийся	std-max	96,7	95,8	94,0	91,1	86,8	81,2
	softmax	97,9	97,4	96,4	94,6	91,8	88,0
	geo-mean	98,2	97,9	97,1	95,6	93,2	90,0
Бурый	std-max	89,8	88,1	85,0	80,3	74,1	66,7
	softmax	92,7	91,8	90,0	87,2	83,1	77,8
	geo-mean	93,2	92,5	91,1	88,8	85,3	80,6

Таблица 8
Точность классификации углей при изменении контрастности, %

Table 8
Accuracy of the coal classification in the contrast variation, %

Марка угля	Пулинг	Коэффициент контрастности					
		0,5	0,7	0,9	1,1	1,3	1,5
Антрацит	std-max	91,4	95,6	97,8	97,3	94,9	90,2
	softmax	95,7	97,9	98,8	98,5	97,4	95,1
	geo-mean	96,8	98,5	99,2	99,0	98,2	96,5
Коксующийся	std-max	88,1	93,2	96,1	95,4	92,0	86,5
	softmax	93,4	96,3	97,6	97,1	95,5	92,6
	geo-mean	94,9	97,2	98,1	97,8	96,6	94,3
Бурый	std-max	79,8	85,7	89,0	87,9	83,5	76,8
	softmax	86,6	90,5	92,2	91,3	88,6	84,2
	geo-mean	88,7	91,9	93,0	92,4	90,3	86,8

Таблица 9
Точность классификации углей при изменении яркости, %

Table 9
Accuracy of the coal classification in the brightness variation, %

Марка угля	Пулинг	Смещение яркости					
		-50	-30	-10	10	30	50
Антрацит	std-max	94,3	97,0	98,1	97,7	95,9	92,8
	softmax	97,2	98,5	99,0	98,8	97,9	96,3
	geo-mean	98,0	98,9	99,3	99,1	98,5	97,4
Коксующийся	std-max	90,6	94,5	96,4	95,8	93,2	89,1
	softmax	94,9	96,8	97,7	97,3	96,0	93,8
	geo-mean	96,1	97,6	98,2	97,9	97,0	95,3
Бурый	std-max	82,9	87,4	89,5	88,7	85,6	81,2
	softmax	88,7	91,3	92,5	91,9	90,0	87,1
	geo-mean	90,5	92,6	93,2	92,7	91,2	88,9

Из табл. 6 и 7 видно, что предложенные методы пулинга обеспечивают более высокую устойчивость к шуму и размытию по сравнению со стандартным максимизирующим пулингом. При этом геометрический пулинг демонстрирует наилучшую robustness, сохраняя точность классификации выше 90% для антрацита и коксующегося угля даже при высоких уровнях искажений. Этот эффект можно объяснить тем, что геометрический пулинг подчеркивает наиболее существенные признаки текстуры углей, игнорируя случайные вариации яркости отдельных пикселей.

Эксперименты с изменением контрастности и яркости (табл. 8 и 9) также показывают преимущество предложенных методов, особенно геометрического пулинга, по сравнению с базовым. Geometric mean лучше сохраняет точность при отклонении контраста и яркости от оптимальных значений за счет более равномерного учета локальных паттернов текстуры угля.

Из табл. 9 видно, что точность классификации для всех марок угля максимальна при небольших отклонениях яркости от исходной (± 10) и снижается при более сильных ва-

Таблица 10
Точность классификации
углей на новых данных, %

Table 10
Accuracy of the coal
classification using new
data, %

Марка угля	Пулинг	Точность классификации	
		Базовый набор	Новые данные
Газовый	std-max	94,5	91,2
	softmax	96,2	94,6
	geo-mean	96,6	95,4
Длиннопламенный	std-max	92,1	88,4
	softmax	94,5	92,7
	geo-mean	94,9	93,5
Бурый	std-max	89,8	85,0
	softmax	92,7	90,2
	geo-mean	93,2	91,3

риациях яркости. При этом геометрический пулинг обеспечивает наибольшую устойчивость к изменениям яркости, а стандартный максимизирующий пулинг – наименьшую. Так, при увеличении яркости на 50 геометрический пулинг обеспечивает точность 97,4% для антрацита, 95,3% для коксующегося и 88,9% для бурого угля, в то время как точность стандартного пулинга составляет лишь 92,8%, 89,1% и 81,2% соответственно. Для этого из каждого набора данных случайным образом отбиралось по 100 изображений, которые затем тестировались моделями, обученными на оставшихся изображениях. Результаты кросс-валидации для наиболее сложных для классификации марок углей (газовый, длиннопламенный и бурый) показаны в табл. 10.

Из табл. 10 видно, что точность классификации на новых данных несколько ниже, чем на основном наборе, что объясняется наличием в новых данных нехарактерных примеров углей, отсутствовавших при обучении. Тем не менее предложенные методы пулинга, особенно геометрический, обеспечивают хорошую обобщающую способность, сохраняя высокую точность классификации на новых данных. Так, для бурого угля геометрический пулинг показывает точность 91,3% на новых данных против 85% у стандартного пулинга. Таким образом, результаты экспериментов на различных марках угля показывают, что предложенные методы пулинга, особенно геометрическое и гармониче-

ское среднее, обеспечивают более высокую точность классификации по сравнению со стандартными методами для всех рассмотренных типов углей. При этом преимущество новых методов возрастает для менее метаморфизованных углей (газовый, длиннопламенный, бурый), классификация которых по визуальным признакам является наиболее сложной.

Заключение

В данной работе были предложены новые методы пулинга с автоматической настройкой параметров для повышения точности классификации угольных пород на изображениях с помощью сверточных нейронных сетей. Softmax пулинг, геометрический и гармонический пулинг, а также пулинг с адаптивным размером окна были протестированы на трех наборах данных (Coal Rock Images, Coal and Rock Images, CoalRockImage5k) и показали превосходство над стандартными методами максимизирующего и усредняющего пулинга, обеспечив точность классификации до 99,3% для антрацита, 98,2% для коксующегося и 93,2% для бурого угля против 98,2%, 96,7% и 89,8% у max пулинга. Преимущество новых методов возрастает для менее метаморфизованных углей (газовый, длиннопламенный, бурый). При этом для достижения точности 95% на антраците геометрическому пулингу требуется около 3000 обучающих изображений против 4500 у max пулинга. Геометрический пулинг показал наилучшую робастность, сохраняя точность выше 90% для антрацита и коксующегося угля даже при высоких уровнях искажений (шум со СКО 50, размытие с радиусом 5). На новых данных геометрический пулинг сохраняет точность 95,4%, 93,5% и 91,3% для газового, длиннопламенного и бурого угля против 91,2%, 88,4% и 85% у max пулинга. Вычислительные эксперименты показали, что точность классификации снижается с уменьшением степени метаморфизма угля – от 99,3% для антрацита до 93,2% для бурого угля у геометрического пулинга. Автоматическая настройка размера окна в size adaptive pooling позволила повысить точность на 1–2% для всех марок угля. При наличии шумов и искажений на изображениях геометрический пулинг сохранил точность выше 97%, 95% и 90% для антрацита, коксующегося и бурого угля при СКО шума 20, радиусе размытия 3, изменении контраста на $\pm 30\%$ и яркости на ± 10 единиц, существенно опередив стандартный max пулинг (95,4%, 92,9% и 83% соответственно). Предложенные методы позволяют сократить требуемый объем обучающей выборки на 30-40% при сохранении целевой точности, что важно ввиду высокой стоимости сбора данных. При тестировании на новых изображениях углей геометрический пулинг обеспечил точность 95,4%, 93,5% и 91,3% для газового, длиннопламенного и бурого угля против 91,2%, 88,4% и 85% у стандартного пулинга.

Список литературы / References

1. Nesteruk S., Agafonova J., Pavlov I., Gerasimov M., Latyshev N., Dimitrov D. et al. MineralImage5k: A benchmark for zero-shot raw mineral visual recognition and description. *Computers & Geosciences*. 2023;178:105414. <https://doi.org/10.1016/j.cageo.2023.105414>
2. Ranzato M.A., Boureau Y.-L., LeCun Y. Sparse feature learning for deep belief networks. In: *Advances in Neural Information Processing Systems 20 – Proceedings of the 2007 Conference*. Vancouver, Canada: Neural Information Processing Systems; 2008. Available at: <https://proceedings.neurips.cc/paper/2007/file/c60d060b946d6dd6145dcbad5c4cc6f-Paper.pdf> (accessed: 21.10.2024).

3. LeCun Y., Boser B., Denker J., Henderson D., Howard R., Hubbard W., Jackel L. Handwritten digit recognition with a back-propagation network. In: Touretzky D. (ed.) *Advances in Neural Information Processing Systems* (NIPS 1989), Denver, CO (Vol. 2). Morgan Kaufmann; 1990. Available at: https://proceedings.neurips.cc/paper_files/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf (accessed: 21.10.2024).
4. Lee C.-Y., Gallagher P.W., Tu Z. *Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree*. arXiv preprint arXiv:1509.08985. 10 October 2015. Available at: <https://arxiv.org/pdf/1509.08985> (accessed: 21.10.2024).
5. Yu D., Wang H., Chen P., Wei Z. Mixed pooling for convolutional neural networks. In: Miao D., Pedrycz W., Slezak D., Peters G., Hu Q., Wang R. (eds.) *Rough Sets and Knowledge Technology*. Springer, Cham; 2014, pp. 364–375. https://doi.org/10.1007/978-3-319-11740-9_34
6. Momeny M., Jahanbakhshi A., Jafarnejhad K., Zhang Y.-D. Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. *Postharvest Biology and Technology*. 2020;166:111204. <https://doi.org/10.1016/j.postharvbio.2020.111204>
7. Zhong S., Wen W., Qin J. *Mix-pooling strategy for attention mechanism*. arXiv preprint arXiv:2208.10322. 22 August 2022. Available at: <https://arxiv.org/pdf/2208.10322v1> (accessed: 21.10.2024).
8. Tong Z., Tanaka G. Hybrid pooling for enhancement of generalization ability in deep convolutional neural networks. *Neurocomputing*. 2019;333:76–85. <https://doi.org/10.1016/j.neucom.2018.12.036>
9. Sermanet P., Chintala S., LeCun Y. *Convolutional neural networks applied to house numbers digit classification*. arXiv:1204.3968. 18 April 2012. <https://doi.org/10.48550/arXiv.1204.3968>
10. Nair V., Hinton G. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010, pp. 807–814. Available at: <https://www.cs.toronto.edu/~hinton/absps/reluICML.pdf> (accessed: 21.10.2024).
11. Ait Skourt B., El Hassani A., Majda A. Mixed-pooling-dropout for convolutional neural network regularization. *Journal of King Saud University – Computer and Information Sciences*. 2022;34(8, Part A):4756–4762. <https://doi.org/10.1016/j.jksuci.2021.05.001>
12. Sun M., Song Z., Jiang X., Pan J., Pang Y. Learning pooling for convolutional neural network. *Neurocomputing*. 2017;224:96–104. <https://doi.org/10.1016/j.neucom.2016.10.049>

Информация об авторах

Козлова Ольга Александровна – инженер-программист, Московский государственный университет имени М.В. Ломоносова, г. Москва, Российская Федерация; <https://orcid.org/0009-0002-8271-4578>; e-mail: olgaalex.kozlova@gmail.com

Китов Виктор Владимирович – кандидат физико-математических наук, старший научный сотрудник, Российский экономический университет имени Г.В. Плеханова; доцент, Московский государственный университет имени М.В. Ломоносова, г. Москва, Российская Федерация; <https://orcid.org/0000-0002-3198-5792>; e-mail: v.v.kitov@yandex.ru

Information about the authors

Olga A. Kozlova – Software Engineer, Lomonosov Moscow State University, Moscow, Russian Federation; <https://orcid.org/0009-0002-8271-4578>; e-mail: olgaalex.kozlova@gmail.com

Victor V. Kitov – Cand. Sci. (Phys. and Math.), Leading Engineer, Plekhanov Russian University of Economics; Associate Professor, Lomonosov Moscow State University, Moscow, Russian Federation; <https://orcid.org/0000-0002-3198-5792>; e-mail: v.v.kitov@yandex.ru

Article info

Received: 19.10.2024

Revised: 27.11.2024

Accepted: 04.12.2024

Информация о статье

Поступила в редакцию: 19.10.2024

Поступила после рецензирования: 27.11.2024

Принята к публикации: 04.12.2024