

УДК 004.932

А.А. Зеленский, М.М. Письменскова, В.В. Воронин

## Алгоритм поиска изображений в виде хэш-функций на основе глубинных нейросетевых технологий

Представлен новый метод, основанный на выделении глобальных функций изображения и вычисления двоичных хэш-кодов. Данный подход обеспечивает пиксельное отображение изображения в хэш-пространство. В работе используется методология глубокого обучения для создания дескриптора со свойствами сохранения подобия и статистической независимости. Основным преимуществом данного подхода в отличие от существующих методов является возможность тонкой настройки поиска для специфических приложений, что позволяет получить более качественный результат поиска по сравнению с другими методами. Структура предложенного метода основана на использовании двух типов нейронных сетей: сверточные нейронные сети для описания изображения и автоэнкодера для отображения особенностей изображения в хэш-пространстве. Проведенные эксперименты показали эффективность при поиске изображений в большой базе данных в сравнении с другими известными методами.

**Ключевые слова:** контекстный поиск изображений, глубокая сверточная нейронная сеть, семантическое хэширование, автоэнкодер.

**doi:** 10.21293/1818-0442-2018-21-3-57-62

Системы контекстного поиска изображений находят широкое применение в современном мире. Наиболее популярными являются поисковые запросы по изображению или по семантическому описанию. Такие подходы применяются в системах управления личными фотоколлекциями, поисковых системах для web-ресурсов, медицинских системах и др. Автоматический анализ больших немаркированных наборов изображений практически невозможен без эффективных алгоритмов поиска. Поиск изображений является одной из наиболее значимых проблем в компьютерном зрении и обработке изображений, которая стремительно набирает популярность из-за постоянно увеличивающегося объема фотографий и видео, получаемых ежедневно. Сложность задачи поиска обусловлена визуальной двусмысленностью естественных образов и семантическим разрывом между представлением изображения на основе функций и пониманием человека того, какие изображения похожи, а какие отличаются.

Основная трудность заключается в понимании, какие характеристики наиболее информативны для определения визуального сходства и построения дескриптора изображения. Несмотря на то, что существует большое количество исследований по этой теме, большинство разработанных дескрипторов предполагают использование Евклидовых и других аналогичных метрик для сравнения характеристик и извлечения наиболее близких изображений, что приводит к низкой эффективности работы метода. Альтернативный способ решения этой проблемы – использовать технологии хэширования данных. Это связано с проблемой сохранения подобия при отображении пространства признаков в компактное хэш-пространство.

В последние годы представлено много исследований, в которых сверточные нейронные сети, первоначально предназначенные для распознавания

образов и классификации изображений, используются для различных задач компьютерного зрения. В работе [1] сверточные нейронные сети используются для многоклассовой классификации, которые особенно эффективны для больших наборов данных изображений. На этапе обучения сверточная нейронная сеть изучает иерархию изображения. Как описано в [2], сверточная нейронная сеть способна представить изображение в виде информативного вектора, который описывает семантический смысл изображения таким образом, что в сочетании с соответствующей метрикой может обеспечить хорошие результаты поиска изображений [3].

Вектор функции, представляемый сверточной нейронной сетью, довольно велик, что накладывает ограничения на удобство использования в реальных информационно-поисковых системах, в частности путем привлечения вычислительно-интенсивных показателей, таких как Евклидова метрика, расстояние Кульбака–Лейблера и т.д. Другая сторона проблемы – это ограниченность вектор-функции, т.е. невозможность использовать всю свою описательную мощность для наборов данных разумного размера.

Цель данной работы – разработка алгоритма эффективного поиска изображений в пространстве хэш-функций на основе глубинных нейросетевых технологий.

Для достижения цели предлагается объединить сверточную нейронную сеть, обученную для описания содержимого изображения с помощью автоэнкодера, предназначенного для преобразования этого описания в хэш-значение. Это делается так, что аналогичные значения метрики Хэмминга связаны с изображениями, имеющими похожее семантическое значение.

На рис. 1 представлен способ получения дескриптора изображения на основе хэширования.

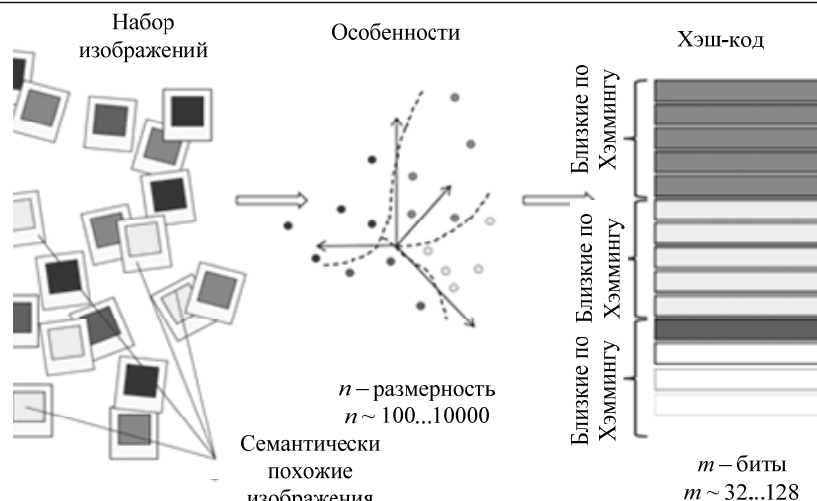


Рис. 1. Дескриптор изображения на основе хэш-функций

### Текущее состояние проблемы

Существующие подходы обучения на основе хэш-функций для поиска изображений можно разделить на три категории: независимые от данных, зависящие от данных и основанные на обучении.

Подходы в первой категории не зависят от размеченных данных и могут рассматриваться как подходы «обучение без учителя» [4–7].

Для второй категории информация о попарных метках используется для изучения хэш-особенностей [8, 9].

Третья категория включает методы, использующие размеченный набор данных в целом для построения процедуры сопоставления [10].

Глубокое обучение нацелено на изучение иерархических особенностей изображений путем создания высокоуровневых особенностей из необработанных данных. Глубокое обучение достигло больших успехов в различных задачах компьютерного зрения, но достаточно небольшой прогресс был сделан в широкомасштабном визуальном поиске на основе хэширования. Семантическое хэширование – это первый подход, в котором для хэширования использовались глубокие методы обучения [11]. Для этой цели была использована сложная ограниченная машина Больцмана (RBM), обученная на создание компактных двоичных кодов. Однако эта модель сложна, требует предварительной подготовки и больших вычислительных мощностей, что ограничивает ее практическое применение. Несмотря на значительное количество литературы по этому вопросу, до сих пор нет надежного и простого в использовании подхода сопоставления изображения и его хэш-значения для использования в системах поиска изображений.

### Бинарное хэширование и функция обучения

Для сравнения «схожести» объектов внутри выбранного класса предлагается получать хэши (компактные двоичные коды) для крупномасштабного визуального поиска, используя методы глубокого обучения. В отличие от существующих подходов обучения двоичных кодов разработана глубокая

нейронная сеть, которая позволяет связывать входное изображение с бинарным вектором по иерархии множественных нелинейных преобразований.

Разработка и обучение данной модели основываются на следующем наборе ограничений:

1. Потеря информации из-за процедуры сопоставления между исходным вектором признаков и вычисленным двоичным кодом должна быть минимизирована.
2. Биты в результирующем двоичном коде равномерно распределены.
3. Дифференциальные биты двоичного кода имеют почти нулевую корреляцию.

Нейронная сеть на основе автоэнкодера удовлетворяет условиям, описанным выше. Результирующая сеть схематически изображена на рис. 2.

Здесь  $I \in \Omega \rightarrow \mathcal{R}$  – входное изображение, где  $\Omega$  – открытое подмножество  $\mathcal{R}^2$  и  $h(I) \in \mathcal{H} = \{0, 1\}^q$  – входное  $q$ -мерное хэш-значение, которое используется для поиска. Учитывая изображение запроса  $I_{\text{query}}$  и набор  $n$ -возможных изображений-кандидатов –  $I_{\text{candidate}} = \{I_p\}$ ,  $p = 1..n$  используются функции, вычисленные сверточной нейронной сетью  $f_{\text{cnn}} \in \mathcal{F} = [0, 1]^{4096}$ , чтобы идентифицировать топ-5 изображений, которые будут формировать набор кандидатов.

Пусть  $f_{\text{cnn}}(I_{\text{query}})$  и  $f_{\text{cnn}}(I_{\text{candidate}})$  являются векторами признаков изображения запроса  $I_{\text{query}}$  и  $I_p$  из набора кандидатов соответственно. Уровень подобия  $S$  между  $I_{\text{query}}$  и  $p$ -изображением из набора кандидатов  $I_{\text{candidate}}$  определяется как евклидово расстояние между соответствующими векторами признаков,

$$s_i = f_{\text{cnn}}(I_{\text{query}}) - f_{\text{cnn}}(I_{\text{candidate}})_2. \quad (1)$$

Для упрощения процедуры обучения и дальнейшего улучшения производительности используется предварительно обученная сеть VGG-19 [12]. Для обучения нейронной сети выбран набор данных CIFAR-100 [11].

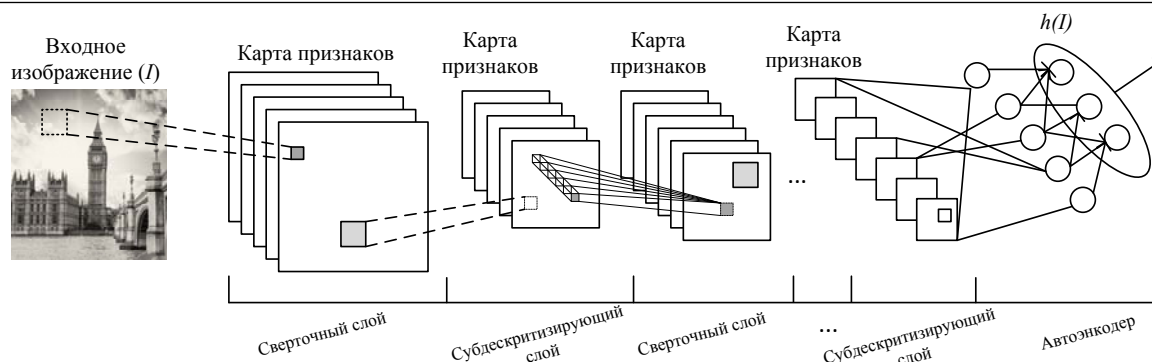


Рис. 2. Дескриптор изображения на основе хэширования

Целевая функция, используемая во время обучения, предназначена для обеспечения максимизации межклассовых вариаций и минимизации внутриклассовых особенностей при вычислении двоичных кодов. Сверточная нейронная сеть использует пространственно-локальную корреляцию, применяя локальный шаблон взаимодействия между нейронами соседних слоев. Кроме того, каждый фильтр нейронной сети реплицируется по всему визуальному полю. Эти реплицированные единицы используют одну и ту же параметризацию (вектор веса и смещение) и образуют карту признаков.

Сеть, с помощью которой вычисляется хэш изображения, состоит из сверточных слоев предобученной сети VGG-19, дескриптора, энкодера, представляющего собой перцептрон, и декодера, структура которого зеркальна структуре энкодера. Входом энкодера служат объединенные признаки сети и дескриптора. Задачей декодера является восстановление информации со входа энкодера. Особенностью автоэнкодера является то, что он сконструирован таким образом, чтобы не иметь возможность точно скопировать вход (в данном случае  $f_{\text{cnn}}(I)$  на выходе. Это достигается за счет меньшего размером скрытого слоя, где и формируется двоичный хэш-код.

Чтобы обеспечить надежные результаты поиска изображений, необходимо построить правильную функцию потерь. Кроме того, совместное использование веса повышает эффективность обучения, значительно уменьшая количество изучаемых свободных параметров. Крупномасштабный поиск изображений на основе контента может быть организован как быстрая процедура поиска в хэш-таблице. В этом случае целью обучения является определение функции  $\phi(f_{\text{cnn}}(I))$ , которая проектирует 4096-размерный дескриптор изображения  $I$  в  $q$ -мерное двоичное хэш-пространство. Проецирование должно быть определено таким образом, чтобы элементы хэш-пространства  $\mathcal{H}$ , назначенные конкретному изображению, сохраняли семантическую согласованность между аналогичными изображениями по расстоянию Хэмминга. Для улучшения результатов на вход нейронной сети подаются данные с верхних слоев сверточной нейронной сети и хэш-код  $h(I)$ , который использует изображение-запрос  $I_{\text{query}}$  как входные данные.

В отличие от данной модели некоторые современные методы предназначены для представления хэш-функции  $h(I)$  как линейной проекции представления изображения. Важная проблема заключается в том, что  $q$ -мерные двоичные метки могут представлять только  $2^q$  разных точек данных. Поскольку число точек данных больше  $2^q$ , многие точки будут отображаться в одну и ту же двоичную метку. Чтобы гарантировать, что точки данных с одной и той же двоичной меткой находятся рядом в пространстве функций, должно выполняться ограничение погрешности квантования:

$$E_q = \frac{1}{n} \sum_{i=1}^n \|x_i - C(x_i)\|, \quad (2)$$

здесь  $n$  — представляет количество обучающих данных;  $x_i$  — обучающий образец;  $C(x_i)$  возвращает центр кластера  $x_i$ .

Во время обучения сети извлекаются функции изображения для размера партии  $n$  (256 в данном случае), набор извлеченных функций  $F = \{f_{\text{cnn}}(I_i)\}$ ,  $i = 1 \dots n$  и значение класса  $\{l_i\}$ ,  $i = 1 \dots n$ . Автоэнкодер принимает в качестве входных данных  $x \in \mathcal{F}$  и обучает параметры  $W = \{\Theta, b\}$  ( $\Theta$  — матрица размером  $4096 \times q$  и  $b$  — смещение). Конечный хэш-код  $\phi(f_{\text{cnn}}(I)) = h(f_{\text{cnn}}(I)) = \sigma(\Theta \times f_{\text{cnn}}(I) + b)$ , где  $\sigma$  является сигмоидной функцией. Восстановление осуществляется следующим образом:

$$\psi(\hat{f}_{\text{cnn}}(I)) = \sigma(\Theta' h(f_{\text{cnn}}(I)) + b'); \quad (3)$$

$$\Theta' = \Theta^T, \quad b' = -b. \quad (4)$$

$$S(i, j) = \begin{cases} 0, & \text{если } x_i \text{ и } x_j \text{ один класс;} \\ 1, & \text{если } x_i \text{ и } x_j \text{ разные классы.} \end{cases} \quad (5)$$

$$NLL = \sum_{i=1}^n \sum_{j=1}^5 S(i, j). \quad (6)$$

Другими словами, автоэнкодер состоит из двух частей энкодер и декодер, которые могут быть определены переходами  $h(f_{\text{cnn}}(I)): \mathcal{F} \rightarrow \mathcal{H}$  и  $h^{-1}(\hat{f}_{\text{cnn}}(I)): \mathcal{H} \rightarrow \mathcal{F}$ . Автоэнкодер обучается минимизировать ошибки восстановления (например, квадратичную ошибку).

$$\mathcal{L}(x, x') = \|x - x'\|^2 = \|x - \sigma(\Theta'(\sigma(\Theta x + b)) + b)\|^2. \quad (7)$$

В данном случае

$$\|f_{\text{cnn}}(I) - h^{-1}(h(f_{\text{cnn}}(I)))\|^2 \rightarrow \min. \quad (8)$$

Если пространство функций  $\mathcal{F}$  имеет меньшую размерность, чем входное пространство  $\mathcal{H}$ , то вектор функции  $\phi(\cdot)$  можно рассматривать как сжатое представление входа  $f_{\text{cnn}}(I)$ . Если скрытые слои автоэнкодера больше входного слоя, то автокодер может потенциально изучить входную функцию и стать бесполезным. Однако экспериментальные результаты показали, что в этих случаях автоэнкодеры могут по-прежнему изучать полезные функции. Регуляризация  $L_1$  и  $L_2$  вводит дополнительную функцию потерь, наложение «штрафа» на определенные параметры. Формально если функцию потерь можно записать как

$$\mathcal{L}(W, F) = \sum_{i=0}^n \mathcal{L}(F_i, \Psi(\phi(F_i))), \quad (9)$$

то регуляризованная функция потерь будет иметь вид

$$\mathcal{L}_{\text{reg}}(W, F) = NLL + \mathcal{L}(W, F) + \lambda \|W\|_p^p, \quad (10)$$

где

$$\|W\|_p = \left( \sum_{i=0}^{|W|} |W_i|^p \right)^{\frac{1}{p}}. \quad (11)$$

В этом выражении  $L_2$  является нормой  $W$ . Параметр  $\lambda$  является гиперпараметром, который контролирует относительную значимость параметра регуляризации. Обычно используемые значения для  $p$  равны 1 и 2, следовательно,  $L_1/L_2$ . Если  $p=2$ , такая регуляризация называется сокращением весов.

Добавление выражения регуляризации к потерям приводит к сглаживанию функций нейронной сети (путем присвоения «штрафов» за большие значения параметров, что уменьшает величину нелинейности). Более интуитивно, два параметра ( $NLL$  и  $\lambda \|W\|_p^p$ ) соответствуют моделированию данных ( $NLL$ ) и имеют «простое» и «гладкое» решение  $\lambda \|W\|_p^p$ .

Изучение оптимальных параметров модели предполагает минимизацию функции потерь. В случае многоклассовой классификации очень часто используют отрицательное логарифмическое правдоподобие как функцию потерь. Это эквивалентно максимизации вероятности  $\mathcal{P}$  набора данных  $F$  в модели, параметризованной  $W$ . Начнем сначала с определения вероятности  $\mathcal{P}$ :

$$\mathcal{P}(W = \{\Theta, b\}, F) = \prod_{i=0}^D \log(P(l = l_i | x_{(i)}, \Theta, b)). \quad (12)$$

Обучение нейронной сети осуществлялось в соответствии с рекомендациями Krizhevsky [1].

А именно, нейронная сеть обучается с использованием многокомпонентной логистической регрессии, мини-пакетного (mini-batch) градиентного спуска (на основе обратного распространения [13]) с импульсом. Размер пакета был установлен на 256, импульс 0,9. Обучение было регуляризовано сокращением весов (множитель штрафа  $L_2$  установлен на 5).

### Экспериментальные результаты

Для проведения экспериментов используется стандартная сеть VGG19 – сверточная нейронная сеть, которая содержит 19 слоев. Сеть состоит из чередующихся каскадов свертки и подвыборки. Сначала идут два каскада свертка-свертка-подвыборка, а затем три каскада свертка-свертка-свертка-подвыборка. Размер узла свертки во всех слоях  $3 \times 3$ . При подвыборке производится выбор максимального значения (Max Pooling) из квадрата размером  $2 \times 2$ .

Описанный подход сравнивался с несколькими современными методами на наборе данных CIFAR-100, который содержит 100 категорий объектов, и каждая категория состоит из 6000 изображений. Таким образом, всего 600000 изображений. Набор данных разделен на обучающие и тестовые наборы, 50000 и 10000 изображений соответственно. Процентное соотношение верхних пяти результатов представлено в таблице, был протестирован каждый подход на хэш-функциях длиной 32, 48, 64 и 128 бит.

Сравнение с известными методами (в %)

Длина хэш-функции (бит)	32	48	64	128
Предложенный метод	<b>45,18</b>	<b>49,34</b>	<b>53,08</b>	66,47
KMH [5]	32,12	33,29	35,78	60,43
Spherical [6]	24,11	30,77	34,75	49,48
SH [7]	27,32	25,72	24,10	56,29
PCAN [9]	28,15	24,85	21,47	56,56
LSH [14]	20,82	25,83	31,71	37,77
DH [15]	43,12	44,97	46,74	<b>66,89</b>
SPLH [16]	43,20	48,29	48,34	62,98

Экспериментальные результаты показывают, что предлагаемый метод способен получать только семантическую информацию, что эффективно для небольших изменений цветовой и текстурной информации. На небольших наборах данных, метод не различает изображения с различной цветовой информацией, но схожим семантическим описанием. Извлечение функций сверточной нейронной сетью занимает около 60 миллисекунд (мс) на ПК с графическим процессором Geforce GTX 860M GPU и объемом памяти 4 Гб. Поиск осуществляется в режиме CPU, реализация на C/C++.

Вычисления Евклидовой метрики между двумя 4096-мерными векторами составляет 109,767 мс. Вычисление расстояния Хэмминга между двумя 128-битными двоичными кодами занимает 0,113 мс. Таким образом, предложенный подход к поиску изображений, основанный на хэшировании, в тысячу раз быстрее, чем традиционный поиск с 4096-мерными функциями.

**Заключение и выводы**

В статье представлен новый метод отображения изображений в пространство хэш-функций. Рассмотренный подход применен к задаче поиска изображений на основе семантической информации. Представлен алгоритм поиска изображений, объединяющий сверточную нейронную сеть и автоэнкодер. Экспериментальные результаты показали, что с помощью модификации глубоких сверточных нейронных сетей можно добиться значимого улучшения точности поиска изображений.

Работа выполнена при поддержке гранта РФФИ №17-57-53192.

**Литература**

1. Krizhevsky A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // *Advances in neural information processing systems*. – 2012. – P. 1097–1105.
2. On the performance of convnet features for place recognition / N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, M. Milford // *Intelligent Robots and Systems (IROS)*, 2015 IEEE/RSJ International Conference. – Hamburg: IEEE, 2015. – P. 4297–4304.
3. Neural codes for image retrieval / A. Babenko, A. Slesarev, A. Chigorin, V. Lempitsky // *European conference on computer vision*. – Cham: Springer, 2014. – P. 584–599.
4. Grauman K. Learning binary hash codes for large-scale image search / K. Grauman, R. Fergus // *Machine learning for computer vision*. – Berlin: Springer, 2013. – P. 49–87.
5. Learning hash codes with listwise supervision / J. Wang, W. Liu, A.X. Sun, Y.G. Jiang // *Proceedings of the IEEE International Conference on Computer Vision*. – 2013. – P. 3032–3039.
6. Li X. Learning hash functions using column generation / X. Li, G. Lin, C. Shen, A.V.D. Hengel, A. Dick [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1303.0339.pdf>, свободный (дата обращения: 25.07.2018).
7. Li J. Fast image search with deep convolutional neural networks and efficient hashing codes / J.Y. Li, J.H. Li // *Fuzzy Systems and Knowledge Discovery (FSKD)*, 2015 12th International Conference on. – Zhangjiajie: IEEE, 2015. – P. 1285–1290.
8. Liong V.E. Deep hashing for compact binary codes learning / V.E. Liong, J. Lu, G. Wang, P. Moulin, J. Zhou // *Proceedings of the IEEE conference on computer vision and pattern recognition*. – 2015. – P. 2475–2483.
9. Asymmetric cyclical hashing for large scale image retrieval / Y. Lv, W.W. Ng, Z. Zeng, D.S. Yeung, P.P. Chan // *IEEE Transactions on Multimedia*. – 2015. – Vol. 17, No. 8. – P. 1225–1235.
10. Packing and padding: Coupled multi-index for accurate image retrieval / L. Zheng, S. Wang, Z. Liu, Q. Tian // *Proceedings of the IEEE conference on computer vision and pattern recognition*. – 2014. – P. 1939–1946.
11. Krizhevsky A. Learning multiple layers of features from tiny images / A. Krizhevsky, G. Hinton // *Technical report*, University of Toronto. – 2009. – Vol. 1, No. 4. – 60 p.
12. Going deeper with convolutions / C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich // *Proceedings of the IEEE conference on computer vision and pattern recognition*. – 2015. – P. 1–9.
13. Caffe: Convolutional architecture for fast feature embedding / Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell // *Proceedings of the 22nd ACM international conference on Multimedia*. – Orlando: ACM, 2014. – P. 675–678.
14. Kulis B. Learning to hash with binary reconstructive embeddings / B. Kulis, T. Darrell // *Advances in neural information processing systems*. – 2009. – P. 1042–1050.
15. Lin K. Deep learning of binary hash codes for fast image retrieval / K. Lin, H.F. Yang, J.H. Hsiao, C.S. Chen // *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. – 2015. – P. 27–35.
16. Zhao H. A fast binary encoding mechanism for approximate nearest neighbor search / H. Zhao, Z. Wang, P. Liu, B. Wu // *Neurocomputing*. – 2016. – Vol. 178. – P. 112–122.

**Зеленский Александр Александрович**

Канд. техн. наук, проректор по научной работе и научно-технической политике  
Московского государственного технологического университета «СТАНКИН»  
Вадковский пер., д. 3а, г. Москва, Россия, 127055  
Тел.: +7 (499) 972-95-21  
Эл. почта: [science@stankin.ru](mailto:science@stankin.ru)

**Письменкова Марина Михайловна**

Аспирант Московского государственного технологического университета «СТАНКИН»  
Вадковский пер., д. 3а, г. Москва, Россия, 127055  
ORCID 0000-0001-9485-4696  
Тел.: +7-928-814 68 51  
Эл. почта: [mpismenskova@mail.ru](mailto:mpismenskova@mail.ru)

**Воронин Вячеслав Владимирович**

Канд. техн. наук, инженер Московского государственного технологического университета «СТАНКИН»  
Вадковский пер., д. 3а, г. Москва, Россия, 127055  
ORCID 0000-0001-8114-6383  
Тел.: +7-988-534-34-59  
Эл. почта: [voronin\\_sl@mail.ru](mailto:voronin_sl@mail.ru)

Zelensky A.A., Pismenskova M.M., Voronin V.V.  
**Algorithm for image retravel in the space of hash functions based on the deep neural network technologies**

This paper addresses the problem by novel technique for simultaneous learning of global image features and binary hash codes. Our approach provide mapping of pixel-based image representation to hash-value space simultaneously trying to save as much of semantic image content as possible. We use deep learning methodology to generate image description with properties of similarity preservation and statistical independence. The main advantage of our approach in contrast to existing is ability to fine-tune retrieval procedure for very specific application which allow us to provide better results in comparison to general techniques. The framework for data-dependent image presented in the paper hashing is based on use two different kinds of neural networks: convolutional neural networks for image description and autoencoder for feature to hash space mapping. Experimental results confirmed that our approach has shown promising results comparing to other state-of-the-art methods.

**Keywords:** content-based image retrieval, deep convolutional neural network, semantic hashing, autoencoder.

**doi:** 10.21293/1818-0442-2018-21-3-57-62

## References

1. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012, pp. 1097–1105.
2. Sünderhauf N., Shirazi S., Dayoub F., Upcroft B., Milford M. On the performance of convnet features for place recognition. *Intelligent Robots and Systems (IROS), 2015 IEEE / RSJ International Conference*, IEEE, 2015, pp. 4297–4304.
3. Babenko A., Slesarev A., Chigorin A., Lempitsky V. Neural codes for image retrieval. *European conference on computer vision*, Springer, 2014, pp. 584–599.
4. Grauman K., Fergus R. Learning binary hash codes for large-scale image search. *Machine learning for computer vision*. Berlin: Springer, 2013, pp. 49–87.
5. Wang J., Liu W., Sun A.X., Jiang Y.G. Learning hash codes with listwise supervision. *Proceedings of the IEEE International Conference on Computer Vision*. 2013, pp. 3032–3039.
6. Li X., Lin G., Lin C., Shen C., Hengel A.V.D., Dick A. Learning hash functions using column generation [Electronic resource] Access mode: <https://arxiv.org/pdf/1303.0339.pdf>, free (accessed: 25 July 2018).
7. Li J., Li J.H. Fast image search with deep convolutional neural networks and efficient hashing codes. *Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on*, IEEE, 2015, pp. 1285–1290.
8. Erin Liong V., Lu J., Wang G., Moulin P., Zhou J. Deep hashing for compact binary codes learning. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, P. 2475–2483.
9. Lv Y., Ng W.W., Zeng Z., Yeung D.S., Chan P.P. Asymmetric cyclical hashing for large scale image retrieval. *IEEE Transactions on Multimedia*, 2015, Vol. 17, no. 8, pp. 1225–1235.
10. Zheng L., Wang S., Liu Z., Tian Q. Packing and padding: Coupled multi-index for accurate image retrieval. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, P. 1939–1946.
11. Krizhevsky A., Hinton G. Learning multiple layers of features from tiny images – Technical report, University of Toronto, 2009, Vol. 1, № 4, 60 p.
12. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Rabinovich A. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
13. Jia Y., Shelhamer E., Donahue J., Karayev S., Long J., Girshick R., Darrell T. Caffe: Convolutional architecture for fast feature embedding. *Proceedings of the 22nd ACM international conference on Multimedia*, ACM, 2014, pp. 675–678.
14. Kulis B., Darrell T. Learning to hash with binary reconstructive embeddings. *Advances in neural information processing systems*, 2009, pp. 1042–1050.
15. Lin K., Yang H.F., Hsiao J.H., Chen C.S. Deep learning of binary hash codes for fast image retrieval. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 27–35.
16. Zhao H., Wang Z., Liu P., Wu B. A fast binary encoding mechanism for approximate nearest neighbor search. *Neurocomputing*, 2016, Vol. 178, pp. 112–122.

**Alexander A. Zelensky**

Ph.D. in Engineering Science, Vice rector  
for Research Work and R&D Politics, Moscow State  
Technological University «STANKIN»  
3a, Vadkovsky Lane, Moscow, Russia, 127055  
Phone: +7 (499) 972-95-21  
Email: science@stankin.ru

**Marina M. Pismenskova**

Post-graduate student, Moscow State Technological  
University «STANKIN»  
3a, Vadkovsky Lane, Moscow, Russia, 127055  
ORCID 0000-0001-9485-4696  
Phone.: +7-928-814-68-51  
Email: mpismenskova@mail.ru

**Viacheslav V. Voronin**

Ph.D. in Engineering Science, Moscow  
State Technological University «STANKIN»  
3a, Vadkovsky Lane, Moscow, Russia, 127055  
ORCID 0000-0001-8114-6383  
Phone: +7-988-534-34-59  
Email: voronin\_sl@mail.ru