

УДК 0048:681.3

Тимофєєва А.Є.

Запорізький національний університет

Кудін О.В.

Запорізький національний університет

Кривохата А.Г.

Запорізький національний університет

Лісняк А.О.

Запорізький національний університет

АВТОМАТИЧНЕ АНОТУВАННЯ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

У цій роботі розглянуто основні методи та моделі машинного навчання, які застосовуються для розв'язання задачі автоматичного анотування зображень. Нині системи, які мають здатність вилучати значення з візуальних даних, дедалі більше розвиваються та використовуються як у наукових колах, так і в промисловості. Одним із важливих напрямів у рамках цієї задачі є розробка автоматичних систем для розуміння візуальних сцен. Розроблено систему автоматичного анотування зображень із використанням загорткової нейронної мережі. На етапі попередньої обробки зображення застосовується кілька підходів до вилучення візуальних ознак. Окреслено напрям подальшого розвитку систем анотації зображень.

Ключові слова: автоматична анотація зображень, згорткова нейронна мережа, машинне навчання, візуальні ознаки, класифікація.

Постановка проблеми. Автоматичне анотування зображень є досить перспективною областю досліджень, оскільки такі системи можуть використовуватись для пошуку зображень та як підсистеми в системах штучного інтелекту. У методах анотування зображень ключову роль відіграють візуальні ознаки, які вилучаються із зображень і використовуються для їх компактного представлення. У загальному випадку використовувані ознаки можна поділити на ознаки кольору, текстури та форми, своєю чергою, розрізняють регіональні (засновані на локальних характеристиках окремих регіонів зображень) і глобальні (обчислювані для всього зображення загалом) представлення зображень. У розглянутих нижче статтях автори використовують низку класифікаторів, які на основі отриманих візуальних ознак формують відповідь у вигляді ключових слів або міток, що описують зображення. У деяких роботах пропонується підхід, за якого можна отримувати зв'язний текст, що описує зображення.

Аналіз останніх досліджень і публікацій. Існує кілька підходів до класифікації методів автоматичного анотування зображень, наприклад, ймовірнісні та детерміновані методи, методи, засновані

на навчанні, та пошукові методи, контрольовані (з учителем), напівконтрольовані та неконтрольовані методи. Проте більшість літературних джерел [1–3] виділяють п'ять категорій: 1) породжувальні методи (англ. Generative model-based approach), які генерують спільний ймовірнісний розподіл ознак зображення та міток; 2) методи найближчих сусідів (англ. Nearest neighbor model-based approach), які припускають, що зображення з такими ознаками мають велику ймовірність розділяти схожі мітки; 3) розрізнявальні (умовні) методи (англ. Discriminative model-based approach), які розглядають завдання анотації як проблему прогнозування класу зображення залежно від ознак; 4) методи, засновані на завершенні міток (англ. Tag completion model-based approach), які можуть не тільки передбачити, але й виправити мітки, які введено помилково; 5) методи глибинного навчання (англ. Deep learning-based approach), які можуть використовувати для прогнозування як необроблені зображення, так і вилучені задалегідь ознаки.

Далі наведено класичні методи машинного навчання, які за загальноприйнятою класифікацією можна зарахувати до породжувальних методів, умовних та методів найближчих сусідів.

Поширеним є використання SVM (англ. Support Vector Machine) в об'єднанні з алгоритмом KNN (англ. k-Nearest Neighbor) [4; 5]. Мета використання SVM у розв'язанні задачі анотування полягає в тому, щоб створити модель, яка передбачає тільки цільове значення екземплярів даних з атрибутами в тестовому наборі [6]. В роботах [7; 8] було використано двохпрохідний метод k-найближчих сусідів (англ. 2-Pass K-Nearest Neighbor), в якому у процесі анотування нового зображення на першому проході методу з кожної семантичної групи вибирається по два найбільш схожих навчальних зображення, які потім об'єднуються в набір зображень. Отриманий набір включає найбільш інформативні навчаючі зображення для прогнозування приналежності ключових слів новому зображенню. На другому проході методу здійснюється оцінка ймовірностей за допомогою зображень тренувального набору даних.

У роботах науковців [9–11] було запропоновано метод CMRM (англ. Cross-Media Relevance Model), в якому зображення сегментуються на області, після чого з областей, розмір яких більше певного порогу, вилучаються вектори ознак. Всі отримані вектори групуються в кластери візуально схожих областей зображень (візуальні слова), а кожне зображення з навчальної вибірки представляється у вигляді набору візуальних і ключових слів. Також у статті «Multiple Bernoulli relevance models for image and video annotation» [12] було показано, що поділ зображення на прямокутні плитки замість використання методів автоматичної сегментації призводить до підвищення точності і повноти анотування. Ця модифікація отримала назву Multiple Bernoulli Relevance Model (MBRM) [13; 14].

Гібридну модель SVM-DMBRM розроблено у роботі «A Hybrid Model for Automatic Image Annotation» [15]. Автори поєднують метод опорних векторів, що виступає в ролі розрізняючої моделі та дискретний метод MBRM – як породжувальна модель. Ймовірності присвоєння ключових слів зображенню, отримані окремо для класифікаційної і генеративної частин методу, лінійно комбінуються, формуючи ранжований список ключових слів.

У роботі «Supervised LDA for Image Annotation» [16] використовуються методи Latent Semantic Analysis (PLSA) та Latent Dirichlet Allocation (LDA), які зазвичай використовуються в задачах обробки текстових даних.

Автори статті «Automatic image annotation using community detection in neighbor images.» [17] пропонують метод виявлення спільних ознак у сусідніх зображеннях. На першому етапі запропонованого способу сусідні зображення, які схожі із зображен-

ням з запиту, вилучаються за допомогою функцій низького рівня. На другому етапі створюється мережа міток сусідніх зображень, з яких обираються ті, що найбільш релевантні для зображення із запиту.

У роботі «A New Baseline for Image Annotation» [18] було запропоновано використовувати метод Joint Equal Contribution (JEC), який полягає в тому, що зображення описується за допомогою 7 глобальних ознак, які обчислено на основі кольору та текстури зображення, нормалізованих таким чином, щоб значення відстаней між парами ознак будь-яких двох зображень знаходилися в діапазоні [0; 1]. При порівнянні двох зображень спочатку окремо обчислюються відстані для кожного типу ознак, після чого отримані значення об'єднуються з рівними вагами. У процесі анотування нового зображення для нього визначається 5 найближчих навчальних зображень, відсортованих за збільшенням відстані.

Використовують графові моделі, які дають змогу дослідити взаємозв'язки між мітками та використовуваними ознаками [19; 20].

Інша група методів належить до категорії глибинного навчання. Так, наприклад, у роботах деяких науковців [21–23] використано згорткові нейронні мережі. Згорткові нейронні мережі (англ. Convolutional neural network, CNN) використовуються для вилучення візуальних ознак зображення, починаючи з таких базових функцій, як контури зображень, а потім поступово створюють більш складні ознаки, які можуть описувати форму та колір. Також згорткові нейронні мережі поєднують з рекурентними нейронними мережами (англ. Recurrent neural network, RNN). Так, у роботах «Automatic Image Annotation Model Using LSTM Approach, Signal & Image Processing», «An Adaptive Recognition Model for Image Annotation» [25; 26] як RNN використана модель LSTM (англ. Long short-term memory), яка відповідає за формування розширеної текстової анотації до зображення. У статті «An Adaptive Recognition Model for Image Annotation» [26] використано адаптивну модель розпізнання (англ. Adaptive recognition model, ARM). ARM складається з адаптивної класифікаційної мережі і нелінійної кореляційної мережі. Для поліпшення точності семантичного анотування зображень у роботах «A semantic approach for automatic image annotation», «Image Annotation by Moments» [27; 28] було застосовано систему, яка поєднує три класифікатора і комбінації деяких методів вилучення ознак зображень: мультикласовий метод опорних векторів, багатошарова нейронна мережа та класифікатори найближчих сусідів об'єднуються разом для класифікації та надання відповідних ключових слів для цього вмісту.

Анотування зображень може також розглядатися як підмодуль системи пошуку зображень. Так, пошук зображення може бути реалізовано за допомогою отримання візуальної інформації (англ. Content-based visual information retrieval, CBVIR) або за допомогою тексту (англ. Text-based visual information retrieval, TBIR). CBIR аналізує візуальний вміст (колір, текстуру, форму тощо) зображень без врахування метаданих зображення (мітки, оточуючі тексти тощо) для його пошуку, а TBIR, навпаки, використовує метадані зображення для пошуку [29].

З аналізу літературних джерел можна дійти висновку, що наявні методи розв'язання задачі автоматичного анотування зображень дають змогу отримувати задовільні результати класифікації при цьому шляхом великої кількості використовуваних методів, є можливість гнучкого налаштування системи анотування. Вилучення ознак зображення – це один із необхідних компонентів системи анотування зображень, адже представлення зображення у вигляді вектору ознак дає змогу збільшити якість класифікації та зменшити час навчання системи.

У літературних джерелах, які наведено в огляді, майже не розглядається задача автоматичного налаштування параметрів алгоритмів машинного навчання, наприклад, за допомогою генетичних методів. Такі гібридні системи можуть стати предметом досліджень у подальших публікаціях.

Постановка завдання. Метою роботи є побудова системи автоматичного анотування зображень різного змісту на основі згорткових нейронних мереж та проведення обчислювальних експериментів із

метою зменшення кількості шарів при задовільній якості анотування.

Виклад основного матеріалу дослідження.

Особливість задачі автоматичного анотування зображень, порівняно зі звичайною класифікацією, полягає в тому, що одне зображення може належати до кількох класів (англ. Multi-Label Classification, MLC). Це треба враховувати під час налаштування будь-якого класифікатора, тобто функція втрат та метрики оцінки якості моделі мають бути налаштовані саме для задачі MLC.

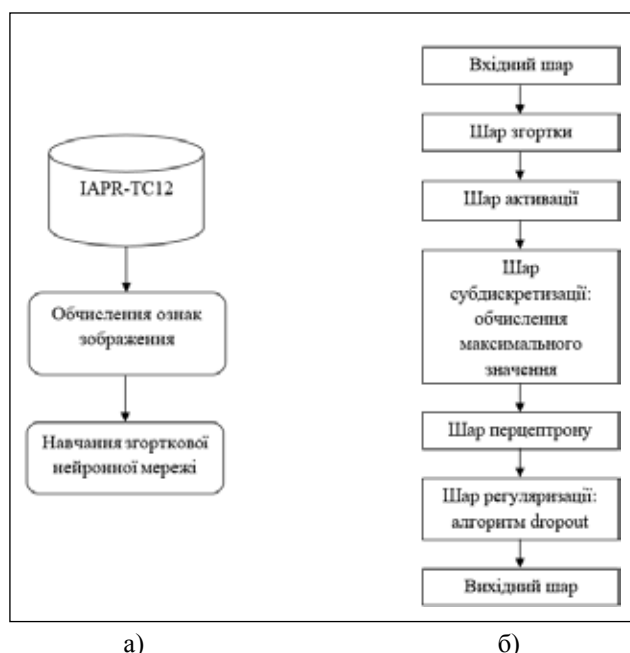
Однією з найбільш застосовуваних архітектур штучних нейронних мереж для задач класифікації зображень є згорткові нейронні мережі. Передбачається використання кількох типів шарів, які чергуються один з одним, утворюючи складну структуру. В шарах згортки виконується афінне перетворення вхідних даних на вихідну мапу ознак. Шари субдискретизації виконують зменшення розмірності мапи ознак, застосовуючи віконну функцію та обираючи максимальне (або середнє) значення серед тих, що потрапили у вікно. У повнозв'язних шарах, на вхід яких поступає набір досить абстрактних ознак, виконується безпосередньо класифікація. Сучасні архітектури згорткових нейронних мереж можуть використовувати десятки шарів, наприклад, мережа VGG-16 (англ. Visual Geometry Group) використовує 16 шарів [30].

Для тренування та тестування нейронної мережі в цій роботі використовується анотований набір даних IAPR-TC12, який складається з 20 000 зображень, із кожним зображенням пов'язаний ієрархічний набір ознак, наприклад, entity->humans->person->child. Докладно набір даних описано в роботі «The segmented and annotated IAPR TC-12 benchmark» [31].

На практиці задовільну якість класифікації засобами згорткових нейронних мереж можна досягнути при невеликій кількості шарів. Зокрема, це досягається шляхом використання попередньої обробки зображень, під час якої обчислюються певні ознаки, наприклад, форми, кольорів тощо. Для зменшення розмірності задачі та пришвидшення обчислень у цій роботі використовуються шість можливих міток для кожного зображення: люди, природа, штучні вироби, тварини, їжа.

Розв'язання задачі автоматичного анотування зображень виконується в кілька етапів (рис. 1а), основне призначення яких – підготовка даних та подальше навчання класифікатора.

У цій роботі пропонується така конфігурація згорткової нейронної мережі: один набір шарів згортки, активації, субдискретизації та два повнозв'язних шари (рис. 1б).



а)
б)
Рис. 1. Етапи розв'язання задачі та структура нейронної мережі

Важливим етапом при навчанні нейронних мереж є визначення багатьох параметрів, які керують процесом навчання, так званих гіперпараметрів мережі. До них належать: спосіб початкової ініціалізації ваг мережі, функції активації та кількості епох навчання. Параметри нейронної мережі та деталі реалізації засобами Python бібліотеки Keras наведено на рисунку 2. Як функція активації використовувалась функція Relu, а на останньому повнозв'язному шарі – функція Sigmoid.

Точність побудованої моделі та функція похибки залежно від епох навчання наведені на рисунку 3. Наводяться значення для навчальної та тестової вибірки.

Висновки. У роботі побудовано систему автоматичного анотування зображень на основі

згорткової нейронної мережі, яка дає змогу поставити у відповідність кожному зображенню кілька текстових міток. У результаті обчислювальних експериментів виявлено, що точність моделі на тестових даних становить приблизно 90%. Варто зазначити, що для розв'язку поставленої задачі достатньою виявилася структура нейронної мережі із одним пакетом шарів згортки-активації-субдискретизації. Перспективи подальших досліджень пов'язані із застосуванням відомих у літературі підходів до побудови ансамблів нейронних мереж та автоматичним налаштуванням гіперпараметрів такої системи методами генетичного програмування.

```
def CreateModel(self):
    batch_size = 10
    data_rows = self.fft_n
    data_cols = self.fft_m
    input_shape = (data_rows, data_cols, 1)
    ##### Визначення шарів згортки та субдискретизації
    self.model = models.Sequential()
    self.model.add(Conv2D(64, (3, 3), input_shape=input_shape, border_mode="same",
        data_format="channels_last", kernel_initializer=self.kernel_initializer,
        bias_initializer=self.bias_initializer, kernel_regularizer=self.kernel_regularizer))
    self.model.add(Activation(self.activation))
    self.model.add(MaxPooling2D(pool_size=(3,3)))
    ##### Визначення шару перцептрону
    self.model.add(Flatten())
    self.model.add(Dense(128, kernel_initializer=self.kernel_initializer, bias_initializer=self.bias_initializer))
    self.model.add(Activation(self.activation))
    self.model.add(Dropout(0.8))
    self.model.add(Dense(6, kernel_initializer=self.kernel_initializer, bias_initializer=self.bias_initializer))
    self.model.add(Activation('sigmoid'))
    ##### Компіляція моделі
    nb_epoch = 100 # Кількість епох навчання
    alpha_zero = 0.0001 # Коефіцієнт швидкості навчання
    optimizer = optimizers.SGD(lr=alpha_zero, momentum=0.0, decay=0.0, nesterov=False)
    self.model.compile(loss = "binary_crossentropy", optimizer=optimizer, metrics = ["binary_accuracy"])
    self.history = self.model.fit(self.xTrain, self.yTrain, batch_size = batch_size,
        epochs = nb_epoch, verbose=2, validation_data = (self.xTest, self.yTest))
```

Рис. 2. Програмна реалізація згорткової нейронної мережі

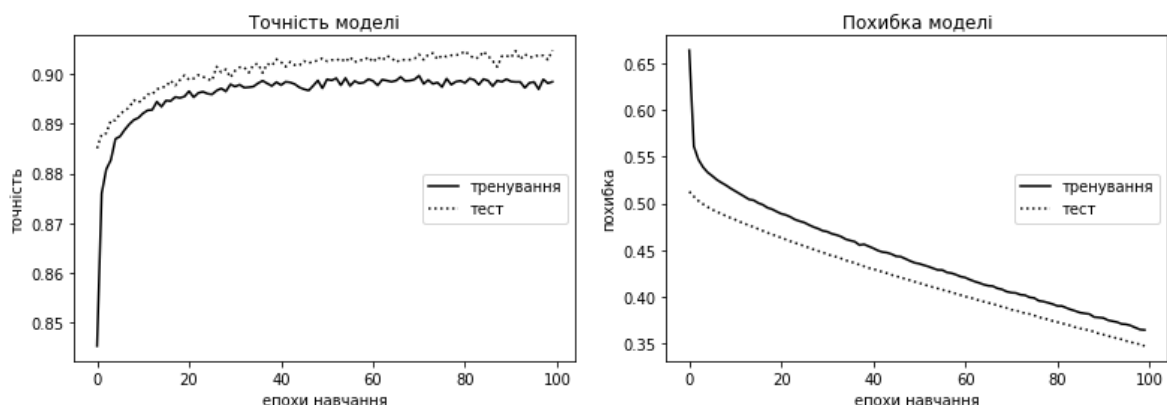


Рис. 3. Точність та похибка розробленої моделі

Список літератури:

1. Qimin Cheng, Qian Zhang, Peng Fu, Conghuan Tu, Sen Li. A survey and analysis on automatic image annotation. *Pattern Recognition* 79. 2018. P. 242–259.
2. Jiwei Hu, Kin-Man Lam, Ping Lou and Quan Liu. Constructing a hierarchical tree for image annotation. *Proceedings of the IEEE International Conference on Multimedia and Expo*. 2017. DOI: 10.1109/ICME.2017.8019512.
3. Bhagata P., Prakash Choudharya. Image Annotation: Then and Now. *Image and Vision Computing*. 2018. doi:10.1016/j.imavis.2018.09.017.
4. Guanglei Chu, Kai Niu, Baoyu Tian. Automatic image annotation combining SVM and KNN algorithm. 3rd International Conference on Cloud Computing and Intelligence Systems. 2014. P. 27–29. DOI: 10.1109/CCIS.2014.71756952014.
5. Cong Jin, Qing-Mei Sun, Shu-Wei Jin. A hybrid automatic image annotation approach. *Multimedia Tools and Applications*. 2018. P. 1–20. URL: <https://doi.org/10.1007/s11042-018-6742-6>.
6. Claudio Cusano, Gianluigi Ciocca. Image annotation using SVM. *University of Milano Bicocca Via Bassini*. 2013. P. 15.
7. Verma Y., Jawahar C.V. Image Annotation Using Metric Learning in Semantic Neighbourhoods. *Proceedings of the 12th European Conference on Computer Vision*. 2012. Vol. 7574. P. 836–849.
8. Artin Ghostan Khatchatoorian, Mansour Jamzad. An Image Annotation Rectifying Method Based on Deep Features. *Proceeding ICDSIP 2018 Proceedings of the 2nd International Conference on Digital Signal Processing*. P. 88–92.
9. Jeon J., Lavrenko V., Manmatha R. Automatic image annotation and retrieval using cross-media relevance models. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*. 2003. Pp. 119–126.
10. Julian Sahertian, Saiful Akbar. Automatic Image Annotation Using CMRM with Scene Information. *Telkomnika*. 2017. Vol. 15, No. 2. Pp. 693–701. DOI: 10.12928/TELKOMNIKA.v15i2.5160.
11. Jeon J., Lavrenko V., Manmatha R. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models, 2003.
12. Feng S.L., Manmatha R., Lavrenko V. Multiple Bernoulli relevance models for image and video annotation. *In Proceedings of the International Conference on Pattern Recognition*. 2004. Vol. 2. Pp. 1002–1009.
13. Ameesh Makadia, Vladimir Pavlovic, Sanjiv Kumar. Baselines for Image Annotation. *Article in International Journal of Computer Vision*, 2010. DOI: 10.1007/s11263-010-0338-6.
14. Feng S.L., Manmatha R. and Lavrenko V. Multiple Bernoulli Relevance Models for Image and Video Annotation. *Multimedia Indexing and Retrieval Group Center for Intelligent Information Retrieval University of Massachusetts*.
15. Venkatesh N. Murthy, Ethem F. Can. A Hybrid Model for Automatic Image Annotation. *Conference Paper*. 2014. DOI: 10.1145/2578726.2578774.
16. Guo Qiaojin, Li Ning, Yang Yubin and Wu Gangshan. Supervised LDA for Image Annotation. *National Key Laboratory for Novel Software Technology Nanjing University Nanjing*.
17. Vafa Maihamia, Farzin Yaghmaee. Automatic image annotation using community detection in neighbor images. *Physica A* 507. 2018. P. 123–132. <https://doi.org/10.1016/j.physa.2018.05.028>.
18. Makadia A., Pavlovic V., Kumar S. A New Baseline for Image Annotation. *Proceedings of the 10th European Conference on Computer Vision*. 2008. Vol. 5304. Pp. 316–329.
19. Jiwei Hu, Kin-Man Lam, Ping Lou, Quan Liu. Constructing a hierarchical tree for image annotation. *The Hong Kong Polytechnic University*. 2017.
20. Shaohua Wan. Image Annotation Using the Simple Decision Tree. *International Conference on Management of e-Commerce and e-Government School of Information and Safety Engineering*. 2011.
21. Utkarsh Ojha, Utsav Adhikari, Dushyant Kumar Singh. Image Annotation Using Deep Learning. *International Conference on Intelligent Computing and Control*. 2017.
22. Ayushi Dutta, Yashaswi Verma-C. V. Jawahar. Automatic image annotation: the quirks and what works. URL: <https://doi.org/10.1007/s11042-018-6247-3>.
23. Yanchun Ma Yongjian, Liu Qing, Xie Lin Li. CNN-feature based automatic image annotation method. URL: <https://doi.org/10.1007/s11042-018-6038-x>.
24. Fangxiang Feng, Ruishan Liu, Xiaole Wang, Hueming Li, Song Bi. Personalized Image Annotation Using Deep Architecture. DOI: 10.1109/ACCESS.2017.2764510
25. Sonu Pratap, Singh Gurjar, Shivam Gupta. Automatic Image Annotation Model Using LSTM Approach, Signal & Image Processing. *An International Journal*. 2017. Vol. 8. No. 4. DOI: 10.5121/sipij.2017.8403.
26. An Adaptive Recognition Model for Image Annotation. *Article in IEEE Transactions on Systems Man and Cybernetics Part*. 2012. DOI: 10.1109/TSMCC.2011.2178831.

27. Mustapha Oujaoura, Brahim Minaoui. M L. A semantic approach for automatic image annotation. *8th International Conference on Intelligent Systems*. 2013. DOI: 10.1109/SITA.2013.6560800.
28. Mustapha Oujaoura, Brahim Minaoui and Mohammed Fakir. Image Annotation by Moments. DOI: 10.15579/gcr.vol11.
29. Karthik Senthil, Abhi Arun and Kamath S. Sowmya. A Content-Based Visual Information Retrieval Approach for Automated Image Annotation. *Chapter in Advances in Intelligent Systems and Computing*. 2018. DOI: 10.1007/978-981-10-3373-5_6.
30. Geron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Sebastopol : O'Reilly. 2017.
31. Hugo Jair Escalante, Carlos A. Hernandez, Jesus A. Gonzalez, Aurelio Lopez, Manuel Montes, Eduardo F. Morales, L. Enrique Sucar, Luis Villaseñor, and Michael Grubinger. The segmented and annotated IAPR TC-12 benchmark. *Computer Vision and Image Understanding*, 114(4):419–428. 2010.

АВТОМАТИЧЕСКОЕ АННОТИРОВАНИЕ ИЗОБРАЖЕНИЙ СРЕДСТВАМИ НЕЙРОННЫХ СЕТЕЙ

В данной работе рассмотрены основные методы и модели машинного обучения, которые применяются для решения задачи автоматического аннотирования изображений. Системы, которые могут извлекать значения из визуальных данных, все больше развиваются и используются как в научных кругах, так и в производстве. Одним из важнейших направлений в рамках этой задачи является разработка автоматических систем для понимания зрительных сцен. Разработана система аннотирования изображений с использованием сверточной нейронной сети. На этапе предобработки изображения извлекаются визуальные признаки. Предложены направления дальнейшего развития систем аннотаций изображений на основе теоретических и экспериментальных моделей.

Ключевые слова: автоматическое аннотирование изображений, сверточная нейронная сеть, машинное обучение, визуальные признаки, классификация.

AUTOMATIC IMAGE ANNOTATION USING NEURAL NETWORKS

This paper discusses the models and methods of machine learning that are employed to solve the problem of automatic image annotation and classification of these methods. Today, the systems which have the ability to extract meaning from visual data are increasingly developed and used both in academia and industry. One of the practically important directions within the scope of this problems is the development of automatic systems for understanding of visual scenes. In this paper, we develop a proof-of-concept system for the image annotation using convolutional neural network. In the image processing stage, we extract visual features. In addition, we outline a direction for further development of image annotating systems based on both theoretical and experimental models.

Key words: Automatic Image Annotation, Convolutional Neural Networks, machine learning, Visual Features, classification.