УДК 004.032.26 doi: 10.15421/322017

Ж.В. Фількінштейн, М.Г. Сидорова
Дніпровський національний університет імені Олеся Гончара

ОСОБЛИВОСТІ ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖЕВИХ МЕТОДІВ ПОШУКУ СХОЖИХ ЗА КОНТЕНТОМ ЗОБРАЖЕНЬ

У роботі проведено дослідження нейромережевих методів пошуку схожих за контентом зображень у поєднанні з сучасними підходами оптимізації процесу навчання моделей згорткових нейронних мереж для виявлення особливостей і факторів, що впливають на якість результатів.

Ключові слова: пошук зображень, моделі згорткових нейронних мереж, глибоке метричне навчання, аугментація, проблема перенавчання.

Z.V. Filkinshtein, M.G. Sydorova
Oles Honchar Dnipro National University

PECULIARITIES OF APPLICATION OF CONTENT BASED IMAGE RETRIEVAL METHODS BASED ON NEURAL NETWORK

This paper investigates the current problem of content based image retrieval on the basis of neural network approach in combination with modern approaches to optimizing the learning process of deep models of convolutional neural networks to identify features and factors influencing the quality of results. Computational schemes have been offered and software has been developed, that allowed to conduct experiments.

The main stages of solving the problem and approaches that can improve the quality of the results have been considered. Comparisons of methods based on categorical crossentropy classification and deep metric learning with triplet loss have been made. The effectiveness of transfer learning approach under its various strategies and ImageNet initialization within the current task has been studied. The problems of overfitting, choosing the architecture of the convolutional neural network, the type of loss function, the method of its optimization and other learning hyper parameters have been considered. Peculiarities and factors influencing the learning process, the quality of results and the specifics of the obtained descriptors have been identified. Several augmentation options (offset, zoom, rotate the image and display horizontally in various combinations) have been also used and its impact has been analyzed.

All research has been conducted in Google Collaborative (cloud service) using proprietary software developed in Python using libraries Keras, TensorFlow, scikit-learn. Visualization of the obtained image descriptors in two-dimensional space has been performed on the basis of the UMAP algorithm. The experiments were performed on different data sets. Convenient visual opportunities for careful analysis of the obtained results have been provided.

Keywords: CBIR (content based image retrieval), deep convolution neural networks, deep metric learning, transfer learning, augmentation, overfitting.

Ж.В. Филькинштейн, М.Г. Сидорова
Днипровський национальний университет имени Олеся Гончара

ОСОБЕННОСТИ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ МЕТОДОВ ПОИСКА ПОХОЖИХ ПО КОНТЕНТУ ИЗОБРАЖЕНИЙ

В работе проведено исследование нейросетевых методов поиска похожих по контенту изображений в сочетании с современными подходами оптимизации процесса обучения моделей сверточных нейронных сетей для выявления особенностей и факторов, влияющих на качество результатов.

Ключевые слова: поиск изображений, модели сверточных нейронных сетей, глубокое метрическое обучение, аугментация, проблема переобучения.

Вступ. У сучасному світі, де кількість графічного контенту стрімко зростає, актуальною є розробка засобів ефективної обробки цих даних. Важливою задачею є пошук схожих за контентом зображень, що може застосовуватися для швидкого доступу, надання рекомендацій, виявлення дублікатів тощо.

Серед існуючих методів розв'язання цієї задачі як і більшості задач аналізу та обробки візуального контенту в останні роки беззаперечним лідером є застосування глибоких згорткових нейронних мереж [2, 5–7]. Проте навчання глибоких моделей є досить складною і нетривіальною задачею, існують проблеми перенавчання, зсуву внутрішніх змінних, невдалої початкової ініціалізації чи вибору інших гіперпараметрів, незбалансованості чи недостатньої кількості даних для навчання.

Метою цієї роботи є дослідження актуального стану проблеми пошуку схожих за контентом зображень на основі нейромережевого підходу в поєднанні з сучасними підходами оптимізації процесу навчання глибоких моделей згорткових нейронних мереж для виявлення особливостей і факторів, що впливають на якість результатів.

Постановка задачі. Запропонувати обчислювальні схеми та розробити програмне забезпечення для здійснення пошуку схожих за контентом зображень та проведення експериментів з метою виявлення впливу функцій втрат та методів їх оптимізації, архітектурних рішень моделей на ефективність отриманих дескрипторів зображень; порівняння підходу глибокого метричного навчання з підходом на основі класифікації; аналізу ефективності застосування transfer learning, засобів запобігання перенавчанню, видів аугментації та її впливу на отримувані результати.

Основні результати. Розглянемо основні етапи розв'язання задачі та підходи, що можуть підвищити якість результатів.

Вибір методу та функції втрат. Головна ідея застосування глибоких згорткових нейронних мереж до обробки зображень полягає у тому, що останні згорткові шари дозволяють виділяти високорівневі характеристики та розглядатися як так звані дескриптори чи ембедінг зображення. Чим ближчими будуть отримані дескриптори у деякому метричному просторі, тим більш схожими будуть зображення, що їм відповідають.

Найбільш відомим методом розв'язання задачі пошуку зображень є підхід на основі класифікації, основна ідея якого полягає у формуванні дескрипторів, шляхом виявлення характерних рис, високорівневих характеристик, що відрізняють об'єкти одного класу від іншого. Для цього останній шар нейронної мережі містить кількість нейронів, що дорівнює кількості класів і, застосовуючи функцію активації SoftMax, інтерпретується як ймовірності приналежності. Основна ідея методу полягає у максимізації ймовірності приналежності кожного вхідного зображення x_i до правильного класу y_i. Для реалізації цієї ідеї в якості функції втрати застосовують функцію Categorical Crossentropy, що розраховується за формулою

$$L = -\frac{1}{N}\sum_{i=1}^{N} \ln p(c = y_i \mid x_i),$$

де c – клас, до якого відноситься зображення за передбаченням; N – кількість зображень [1].

Проте такий метод є дуже вразливим до даних з великою кількістю класів, особливо коли вони є незбалансованими.

Альтернативним підходом до пошуку схожих за контентом зображень є глибоке метричне навчання [3], що також передбачає застосування згорткових нейронних мереж для формування дескрипторів, але на відміну від класифікації, робить це не на основі співставлення класів, а на вивченні принципу схожості і несхожості зображень навчальної вибірки.

Для глибокого метричного підходу останнім шаром є саме дескриптори зображень. До мережі подаються міні-батчі, що складаються з трійок. Трійка являє собою наступний набір:

1. Якір – сформований дескриптор зображення (a).
2. Негатив – сформований дескриптор зображення того ж класу, що і якір (n).
3. Позитив – сформований дескриптор зображення іншого класу (p).

Для експериментів у цій роботі обрано підхід вибору трійок semi-hard triplets, тобто трійки, де негатив не знаходиться ближче до якору ніж позитив, але існує додатня втрата, тобто виконується нерівність

$$d(a, p) < d(a, n) < d(a, p) + margin,$$

де d – функція відстані.

Функцією втрати є функція Triplet, яка формує дескриптори, що відповідають таким ключовим вимогам:

- Два зображення з одного класу мають дескриптори близько розташовані один до одного у метричному просторі.
- Два зображення з різних класів мають дескриптори, розташовані далеко один від одного.

Функцію втрати Triplet розраховують за формулою

$$L = \max(d(a, p) - d(a, n) + margin, 0).$$

Застосування підходу transfer learning. Навчання глибоких згорткових нейронних мереж потребує значних обчислювальних потужностей та затрат часу. За останні роки з'явилася велика кількість моделей створених і навчених з використанням великої кількості даних і великих обчислювальних потужностей. Багато з цих моделей знаходяться у відкритому доступі і можуть бути застосовані для розв'язання нових завдань з метою зменшення часу та підвищення якості навчання. Мета transfer learning – перенести попередньо набуті знання на поточну задачу. Проведені експерименти показали, що застосування початкової ініціалізації вагових коефіцієнтів попередньо навчених на базі даних ImageNet суттєво покращують результати, а вибір стратегії transfer learning залежить від об'єму даних для навчання та його схожості з ImageNet.

Вибір архітектури згорткової нейронної мережі та гіперпараметрів навчання. Рішення стосовно вибору архітектури мережі може суттєвим чином впливати як на якість отримуваних результатів, так і на швидкість навчання. У роботі порівнювалися такі відомі архітектури як VGG, Inception та ResNet. При проведених експериментах, враховуючи час навчання та здатність формувати якісні дескриптори, перевага була надана мережі VGG, її і буде розглянуто надалі.

Не менш важливим питанням є вибір оптимізатора та гіперпараметрів, таких як кількість епох, розмір батчів, швидкість навчання. Розглядалися такі методи як SGD, метод моментів, Adagrad, RMSProp, Adam. У більшості випадків Adam демонстрував найкращі результати, крім того цей метод є найменш чутливим до вибору параметра швидкості навчання.

Усі дослідження проводилися у Google Colaboratory (хмарний сервіс) за допомогою авторського програмного забезпечення, розробленого мовою Python із застосуванням бібліотек Keras, TensorFlow, scikit-learn. Архітектури, що застосовувались для метричного навчання та звичайної класифікаціїї представлені на рис. 1 та на рис. 2 відповідно.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         [(None, 32, 32, 3)]       0
_________________________________________________________________
vgg16 (Functional)           (None, 1, 1, 512)         14714688
_________________________________________________________________
flatten (Flatten)            (None, 512)               0
_________________________________________________________________
dense (Dense)                (None, 512)               262656
_________________________________________________________________
dropout (Dropout)            (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 512)               262656
_________________________________________________________________
lambda (Lambda)              (None, 512)               0
=================================================================
Total params: 15,240,000
Trainable params: 525,312
Non-trainable params: 14,714,688
_________________________________________________________________
```

Рис. 1. Модель, що використовує функцію втрати Triplet

Layer (type)	Output Shape	Param #
input_4 (InputLayer)	[(None, 32, 32, 3)]	0
vgg16 (Functional)	(None, 1, 1, 512)	14714688
flatten_1 (Flatten)	(None, 512)	0
dense_2 (Dense)	(None, 512)	262656
dropout_1 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 512)	262656
lambda_1 (Lambda)	(None, 512)	0
dense_4 (Dense)	(None, 10)	5130

Total params: 15,245,130
Trainable params: 530,442
Non-trainable params: 14,714,688

Рис. 2. Модель, що використовує функцією втрати Categorical Crossentropy

Експерименти проводилися на таких наборах даних: Cifar10, Food, Flowers, Natural, Paris. У результаті експериментів було отримано відображення дескрипторів зображень у двовимірному просторі (рис. 3) за допомогою UMAP, що дозволяє зменшити розмірність з мінімальною втратою інформації. Таке відображення призводить до заключення, що використання моделі з функцією втрати Triplet сприяють більшому зближенню дескрипторів зображення один до одного у метричному просторі ніж використання функції втрати Categorical Crossentropy. Це пояснюється тим, що моделі, які використовують Triplet функції, мають на меті зблизити схожі вихідні вектори та віддалити дескриптори зображень, що належать різним класам. Модель з функцією втрати Categorical Crossentropy спрямована безпосередньо на класифікацію.

Також варто зазначити, що використання шару, з якого отримуються дескриптори, без функції активаціїї та з додаванням регуляризатору L2 призводить до більшого відштовхування дескрипторів зображень, що належать до різних класів, що продемонстровано на рис. 3.

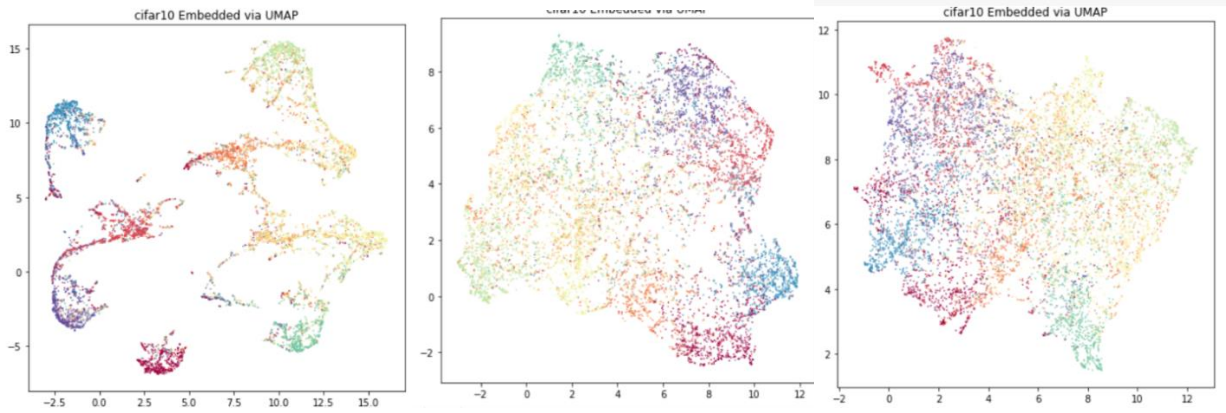

Рис. 3. Відображення дескрипторів зображень, отриманих в результаті моделі з втратою Triplet, моделі з втратою Categorical Crossentropy, моделі з втратою Categorical Crossentropy (з функцією активації RELU на передостанньому шарі, та без регуляризатора L2) відповідно

Було досліджено матрицю сплутаності, за якої можна виявити класи, які найтяжче відлічити один від одного для нейронної мережі. Приклад матриці сплутаності продемонстрований на рис. 4. У випадку з набором даних Cifar10 в усіх проведених експериментах при вхідному зображенні, що відноситься до класу вантажівки, найчастішою помилкою при пошуку схожих зображень були ті, що відносяться до класу автомобіля. Також присутній зворотній випадок: коли при вхідному зображенні автомобіля були отримані помилкові зображення, що відносяться до класу вантажівки, але кількість таких випадків є меншою. Аналогічною ситуацією є сплутування зображень, які входять до класу кішки та собаки. Сплутування деяких класів призводить до погіршення якості навчання моделі в цілому. В даних експериментах було отримано перенавчання (overfitting) моделі, тобто втрату здатності моделі узагальнювати дані.

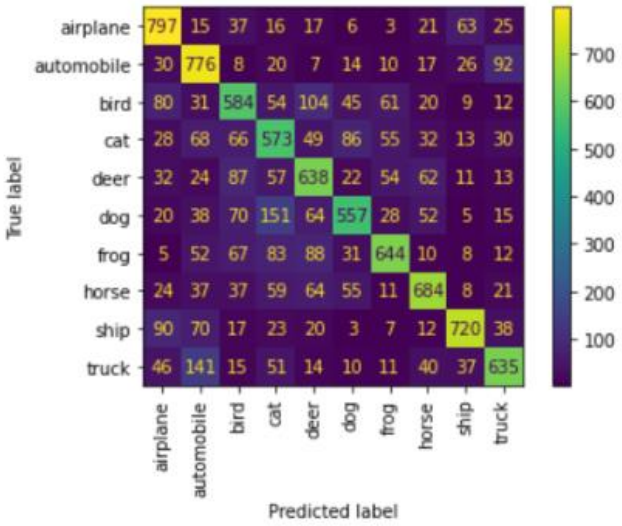

Рис. 4. Матриця сплутаності

Розв'язання проблеми перенавчання (overfitting). Актуальною проблемою при навчанні нейронних мереж є перенавчання моделі, тобто втрата здатності моделі узагальнювати дані, шляхом занадто значного прилаштування до тренувальних даних. Існують різні методи боротьби з перенавчанням: рання зупинка, dropout, регуляризація та інші. В якості засобу, що дозволило моделі зменшити перенавчання, був доданий dropout. Dropout виключає на кожній ітерації випадковим чином нейрони з навчання з ймовірністю p, тобто ймовірність того, що нейрон залишиться в мережі складає (1-p). Але Dropout не усунув проблеми перенавчання. Тому було застосовано ще один метод – регуляризація, яка не дозволяє ваговим коефіцієнтам зростати неконтрольовано. Проте у випадку, коли об'єм даних для навчання є недостатнім, дієвим способом є застосування аугментації.

Аугментація є технікою створення додаткових навчальних даних з наявних шляхом їх деякого перетворення з метою збільшення об'єму навчальної вибірки та збільшення її варіативності [4, 8]. У роботі було застосовано 2 методи аугментації:

1. Зсув, масштабування та поворот зображення (рис. 5).

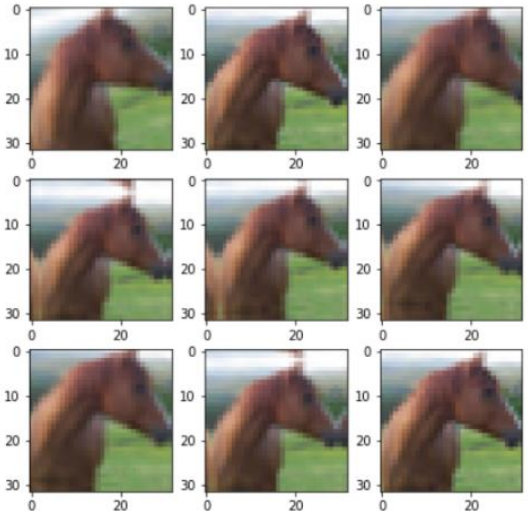

Рис. 5. Приклади зсуву, масштабування та повороту зображення

2. Зсув, масштабування, поворот зображення та горизонтальне відображення. Приклади трансформування представлені на рис. 6.

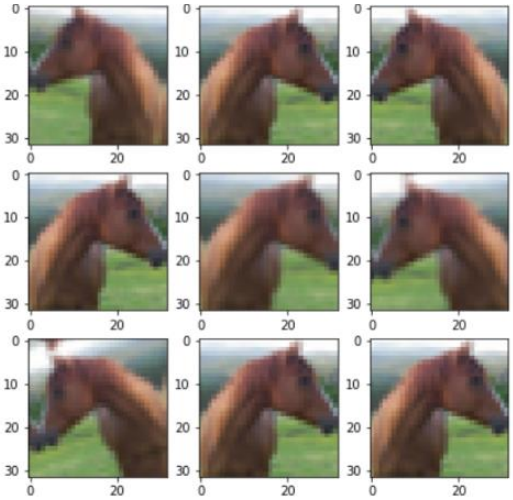

Рис. 6. Приклади зсуву, масштабування, повороту та горизонтального відображення зображення

З результатів в таблицях 1–2 та на рис. 7–8 видно, що аугментація сприяє зменшенню перенавчання, але також може мати і негативний вплив. Через спробу узагальнити дані, отримано, що втрата стала зменшуватись повільніше на тренувальному наборі даних з кожною епохою та якість навчання зменшилась. Також варто зауважити, що модель, що використовує за функцію втрати Triplet є більш стійкою до перенавчання, ніж модель, що використовує функцію втрати Categorical Crossentropy.

Таблиця 1

Результати отримані за моделі з функцією втрати Categorical Crossentropy

	модель з функцією помилки Categorical Crossentropy без аугментації	модель з функцією помилки Categorical Crossentropy з застосування 1-го методу аугментації	модель з функцією помилки Categorical Crossentropy з застосування 2-го методу аугментації
KNN Score	0.6948	0.6845	0.6906
Min validation loss	1.1055	1.0812	1.0708
Epoch with min validation loss	9	21	27
Min train loss	0.7587	0.9846	0.9946
Epoch with min train loss	19	31	37
s/ epoch	16	18	21

Таблиця 2

Результати отримані за моделі з функцією втрати Triplet

	модель з функцією помилки Triplet без аугментації	модель з функцією помилки Triplet з застосування 1-го методу аугментації	модель з функцією помилки Triplet з застосування 2-го методу аугментації
KNN Score	0.6691	0.6697	0.6672
Min validation loss	0.7313	0.7385	0.7352
Epoch with min validation loss	37	44	48
Min train loss	0.5531	0.7189	0.7286
Epoch with min train loss	50	47	49
s/ epoch	16	16	24

Рис. 7. Графіки навчання моделі з функцією втрати Categorical Crossentropy за відсутності аугментації, застосування 1-го методу та 2-го методу аугментації відповідно

Рис. 8. Графіки навчання моделі з функцією втрати Triplet за відсутності аугментації, застосування 1-го методу та 2-го методу аугментації відповідно

Результати пошуку схожих зображень продемонстровані на рис. 9 – 11. При вхідному зображенні вантажівки були отримані різні схожі зображення в залежності від навченої моделі. Так як дескриптори, отримані з моделі з функцією втрати Categorical Crossentropy є більш скупченими, то можливо отримати меш схожі зображення на виході, як у випадку зображень птаха чи жаби, що є менш схожими з вантажівкою ніж зображення автомобіля.

Рис. 9. Вхідне зображення

Рис. 10. Отримані результаті за моделі з функцією втрати Categorical Crossentropy

Рис. 11. Отримані результаті за моделі з функцією втрати Triplet

Висновки. У результаті роботи було запропоновано обчислювальні схеми та розроблено програмне забезпечення, на основі якого проведено дослідження нейромережевих методів пошуку схожих за контентом зображень у поєднанні з сучасними підходами оптимізації процесу навчання моделей згорткових нейронних мереж (transfer learning, dropout, регуляризація, батч-нормалізація, аугментація, різні оптимізатори, архітектури та налаштування гіперпараметрів). Було виявлено особливості і фактори, що впливають на процес навчання, якість результатів та специфіку отримуваних дескрипторів. Здійснено порівняння підходів на основі класифікації та глибокого метричного навчання. Візуалізація отриманих дескрипторів зображень у двовимірному просторі була здійснена на основі алгоритму UMAP.

Подальшими дослідженнями можуть бути експерименти з застосуванням інших технік аугментації, використання вагових коефіцієнтів до класів для розрахунку помилки. Також можливі подальші дослідження на інших наборах даних та за різних архітектур.

Бібліографічні посилання

1. **Goodfellow, I.** Deep Learning. [Text] / I. Goodfellow, Y. Bengio, A. Courville. – MIT Press, 2016. – 775 p.
2. **He, K.** Deep Residual Learning for Image Recognition [Text] / K. He, X. Zhang, Sh. Ren, J. Sun // Computer Vision and Pattern Recognition. – 2016. – №1. – P. 770–778.
3. **Kaya, M.** Deep Metric Learning [Text] / M. Kaya, H. Bilge // Symmetry,– 2019. – №11(9) – P. 1066:1–26
4. **Kumar, H.** Data augmentation Techniques [Електронний ресурс] – Режим доступу: https://iq.opengenus.org/data-augmentation/
5. **Saritha, R.** Content based image retrieval using deep learning process [Text] / R. Saritha, V. Paul, P. Kumar // Cluster Computing. – 2019. – №22. – P. 4187–4200.
6. **Tan, M.** EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [Text] / M. Tan, Q. V. Le. // Proceedings of the 36th International Conference on Machine Learning, Long Beach. – 2019. – P. 6105-6114.
7. **Tzelepi, M.** Deep convolutional learning for Content Based Image Retrieval [Text] / M. Tzelepi, A.Tefas // Neurocomputing. – 2018. – №275. – P. 2467–2478.
8. **Wang, L.** Learning Two-Branch Neural Networks for Image-Text Matching Tasks [Text] / L. Wang, Y. Li, J. Huang, S. Lazebnik // Computer Vision and Pattern Recognition. – 2018. – №1. – P. 1–14.

Надійшла до редколегії 15.10. 2020.