УДК 004.896
DOI https://doi.org/10.32838/2663-5941/2019.6-1/22

Олексій А.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Корнага Я.І.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ІНТЕЛЕКТУАЛЬНА СИСТЕМА РОЗПІЗНАВАННЯ ВІЗУАЛЬНИХ КЛЮЧІВ ДЛЯ МОБІЛЬНИХ ДОДАТКІВ

У світі постійно зростає кількість даних про зображення та сам темп цього зросту. У 2019 році понад 1,8 мільярда зображень щодня завантажували найпопулярніші платформи, такі як Instagram та Facebook. Щодо відеоданих, то, згідно з інформацією компанії Google, на сервіс YouTube кожного дня завантажується більше 86 тисяч годин відео. У світі є безліч девайсів різних типів, що можуть робити фото- та відеозйомку. Персональні смартфони та інші девайси мають змогу створювати знімки з великим розширенням; фототехніка доступна як ніколи. Також цілком доступним стає відеоустаткування з функцією автоматичного захоплення зображень. Автомобілі устатковуються камерами, що дають змогу ефективно слідкувати за трафіком і запобігати нещасним випадкам. Щоб ефективно керувати всіма цими даними, нам потрібно мати деяке уявлення про їхній зміст, а саме – про об'єкти та деталі, що присутні на зображенні. Автоматизована обробка вмісту зображення корисна для широкого різноманіття завдань, пов'язаних із зображеннями. Для комп'ютерних систем це означає схрещування так званого семантичного розриву між інформацією про пікселі, що зберігається у файлі, та розумінням інформаційного наповнення зображення. Автоматична обробка дає можливість ідентифікувати об'єкти, що містяться у файлах зображень, опрацювати та описати деталі та параметри цих об'єктів. Це називається виявленням об'єктів і є однією з основних проблем комп'ютерного зору. Автоматичним системам і роботам є критично важливою можливість оцінювати візуальну сцену навколо себе та мати змогу аналізувати простір навколо себе, опрацьовувати інформацію. Як ми продемонструємо, нині згорткові нейронні мережі є сучасним рішенням для виявлення об'єктів. У статті буде розкрито роботу інтелектуальної системи розпізнавання візуальних ключів для мобільних додатків, а саме – виявлення об'єктів на зображеннях за допомогою згорткових нейронних мереж. Також буде проведено аналіз роботи нейронних систем, виявлено, як можна навчити цю мережу розпізнавати об'єкти, потрібні для процесу верифікації, як можна покращити роботу цього алгоритму за допомогою самонавчання нейронної мережі, питання побудови інфраструктури цієї системи та які можуть бути помилки роботи нейронної мережі у разі помилок. Загалом стаття має розкрити питання призначення згорткової нейронної мережі та доцільність її використання.

Ключові слова: *згорткова нейронна мережа,інтелектуальна система, машинне навчання, алгоритм, комп'ютерний зір.*

Постановка проблеми. Розпізнавання образів є однією з найфундаментальніших проблем теорії інтелектуальних систем. З іншого боку, завдання розпізнавання образів має величезне практичне значення. Замість терміна «розпізнавання» часто використовується інший термін – «класифікація». Ці два терміни в багатьох випадках розглядаються як синоніми, але не є повністю взаємозамінюваними. Кожний із цих термінів має свої сфери застосування, і інтерпретація обох термінів часто залежить від специфіки конкретного завдання.

Завдання ідентифікації, яке полягає в тому, щоби вирізнити певний конкретний об'єкт серед йому подібних (наприклад, впізнати серед інших людей свою дружину), – віднесення об'єкта до того чи іншого класу. Це може бути, наприклад, завдання розпізнавання літер або прийняття рішення про наявність дефекту в деякій технічній деталі. Віднесення об'єкта до певного класу відображає найтиповішу проблему класифікації, і, коли говорять про розпізнавання образів, найчастіше мають на увазі саме цю проблему. Саме вона розглядається тут у першу чергу. Кластерний аналіз полягає в розділенні заданого набору об'єктів на класи – групи об'єктів, схожі між собою за тим чи іншим критерієм. Це завдання часто називають

класифікацією без учителя, оскільки, на відміну від завдання 2, класи апріорно не задані.

Проблеми розпізнавання легко вирішуються людьми, причому робиться це, як правило, підсвідомо. Однак спроби побудувати штучні системи розпізнавання не настільки переконливі. Основна проблема полягає в тому, що здебільшого неможливо адекватно визначити ознаки, на основі яких слід здійснювати розпізнавання. Для завдань, для яких такі ознаки вдається виділити, штучні системи розпізнавання набули значного поширення і широко використовуються.

У випадку цієї статті основним завданням є ідентифікація об'єкта візуального ключа на зображенні, тобто вирішення завдання класифікації.

Аналіз останніх досліджень і публікацій. Проаналізувавши наявні рішення для розпізнавання об'єктів, на зображеннях було виявлено велику кількість підходів та інтелектуальних систем, що вирішують це питання. Наприклад, ця технологія широко використовується в AR-додатках для побудови взаємодії між середовищем, що знімає камера, і програмним забезпеченням. Хорошим прикладом є відома мобільна гра Pokemon Go. У цьому програмному продукті використовується система потокового аналізу відеоконтенту, що використовується для побудови ігрового досвіду в доповненій реальності. Система знаходить та аналізує об'єкти, що дає можливість використовувати їх з ігровою метою. Проте така система дає багато хибних результатів, тому є низка компаній, що вдосконалюють свої інтелектуальні системи та покращують розпізнавання об'єктів. Важливо додати, що можливість побудови доступу до цієї системи через протокол REST є важливим елементом архітектури в контексті сучасних вимог до інтелектуальних систем. Власне, саме реалізація доступу через API дає змогу використовувати цю систему з мобільними додатками.

Постановка завдання. Інтелектуальна система розпізнавання візуальних ключів для мобільних додатків складатиметься з додатка, який буде оброблювати вхідні фото для сканування та знаходження об'єкта візуального ключа. Головне завдання нейронної мережі – знаходження конкретного об'єкта на фото.

Виклад основного матеріалу дослідження. Комп'ютерне бачення займається вилученням змістовної інформації із вмісту цифрових зображень чи відео. Це відрізняється від простої обробки зображень, яка передбачає маніпулювання візуальною інформацією на рівні пікселів.

Застосування комп'ютерного зору включає класифікацію зображень, візуальне виявлення, реконструкцію 3D-сцени з двовимірних зображень, пошук зображень, доповнену реальність, машинне бачення та автоматизацію руху.

Сьогодні машинне навчання є необхідним складником багатьох алгоритмів комп'ютерного зору. Такі алгоритми можна описати як поєднання обробки зображень і машинного навчання. Для ефективних рішень потрібні алгоритми, які можуть впоратися з величезною кількістю інформації, що міститься у візуальних зображеннях, і критично для багатьох застосувань можуть здійснювати обчислення в режимі реального часу.

Виявлення об'єктів є однією з класичних проблем комп'ютерного зору і часто описується як складне завдання. Багато в чому він схожий з іншими завданнями комп'ютерного зору, оскільки передбачає створення рішення, інваріантного деформації та зміні освітлення й кута зору. Що робить виявлення об'єктів серйозною проблемою, то це те, що воно включає як локалізацію, так і класифікацію ділянок зображення. Розмішувальна частина не потрібна, наприклад, у всій класифікації зображень.

Щоб виявити об'єкт, нам потрібно мати уявлення про те, де може бути об'єкт і як зображення сегментується. Це створює тип куряче-яєчної проблеми, де для розпізнавання форми (та класу) предмета нам потрібно знати його розташування, а для розпізнавання місця розташування предмета, нам потрібно знати його форму. Деякі види, що відрізняються візуально, такі як одяг і обличчя людини, можуть бути частинами одного предмета, але це важко пізнати, не розпізнавши спочатку предмет. З іншого боку, деякі об'єкти виділяються лише трохи від фону, вимагаючи розділення перед розпізнаванням.

Важливою проблемою в питанні реалізації комп'ютерного зору за допомогою традиційних нейронних мереж є те, що навіть просте зображення має величезну кількість інформації. Монохромне зображення 620x480 містить 297 600 пікселів. Якщо інтенсивність кожного пікселя цього зображення вводиться окремо до повністю пов'язаної мережі, кожному нейрону потрібно 297 600 ваг. Повне HD-зображення 1920x1080 вимагатиме 2 073 600 ваг. Якщо зображення поліхромні, кількість ваг множиться на кількість кольорових каналів (як правило, три). Отже, ми можемо бачити, що загальна кількість вільних параметрів у мережі швидко стає надзвичайно великою, оскільки розмір зображення збільшується. Занадто великі моделі викликають недоопрацювання та повільну продуктивність.

Крім того, багато завдань на виявлення шаблонів вимагають, щоб рішення було інваріантним для перекладу. Неефективно тренувати нейрони окремо розпізнавати ту саму схему в лівому верхньому куті та в правому нижньому куті зображення. Повністю пов'язана нейронна мережа не може прийняти такого роду структуру до уваги.

Згорткові нейронні мережі є одним з якісних рішень для вирішення цих проблем. Основна ідея CNN (convolutional neural networks) була натхненна біологічною концепцією під назвою рецептивне поле. Рецептивні поля – особливість зорової кори тварин. Вони виконують роль детекторів, чутливих до певних видів подразників, наприклад кутів. Вони розміщені поперек зорового поля та перекриваються один з одним.

Цю біологічну функцію можна реалізувати в комп'ютерах за допомогою операції згортання. Під час обробки зображень зображення можна фільтрувати за допомогою згортки для отримання різних видимих ефектів. На рисунку 1 показано, як обраний вручну згортковий фільтр виявляє горизонтальні краї зображення, функціонуючи аналогічно до рецептивного поля.

Насправді, крапковий добуток фільтра g та під-зображення f (з тими ж розмірами, як g), орієнтовані на координати x, y, створюють значення пікселя h за координатами x, y. Розмір приймального поля регулюється розміром матриці фільтра. Вирівнювання фільтра послідовно з кожним під-зображенням f створює вихідну піксельну матрицю h. У разі нейронних мереж матрицю виводу також називають картою функцій (або картою активації після обчислення функції активації). До країв потрібно ставитися як до окремої справи. Якщо зображення f не зафіксовано, розмір виводу з кожним згортком трохи зменшується.

Набір згорткових фільтрів можна комбінувати для формування згорткового шару нейронної мережі. Матричні значення фільтрів розглядаються як параметри нейронів і навчаються за допомогою машинного навчання. Операція згортки замінює операцію множення звичайного шару нейронної мережі. Вихід шару зазвичай описується як об'єм. Висота і ширина об'єму залежать від розмірів карти активації. Глибина об'єму залежить від кількості фільтрів. Оскільки однакові фільтри використовуються для всіх частин зображення, кількість вільних параметрів різко зменшується порівняно з повністю пов'язаним нейронним шаром. Нейрони згорткового шару здебільшого мають однакові параметри і з'єднані лише з локальною ділянкою входу. Обмін параметрами в результаті згортання забезпечує інваріантність перекладу. Альтернативний спосіб опису звивистого шару – це уявити повністю пов'язаний шар із нескінченно сильним попереднім розміщенням на його вагах. Це примушує нейрони ділити ваги в різних просторових місцях і мати нульову вагу поза полем сприйняття.

Послідовні згорткові шари (часто поєднуються з іншими типами шарів, наприклад, об'єднання, описане нижче) утворюють згорткову нейронну мережу (CNN). Приклад згорткової мережі показано на рисунку 2. Теоретично шари, ближчі до вводу, повинні навчитися розпізнавати особливості зображення на низькому рівні, такі як краї та кути, а шари, наближені до виводу, повинні навчитися поєднувати ці особливості для розпізнавання більш значущих форм. У статті з'ясовано питання того, як можуть згорткові мережі навчитися розпізнавати цілісні об'єкти.

Щоб зробити мережу більш керованою для класифікації, корисно зменшити розмір карти активації в глибокому кінці мережі. Взагалі,

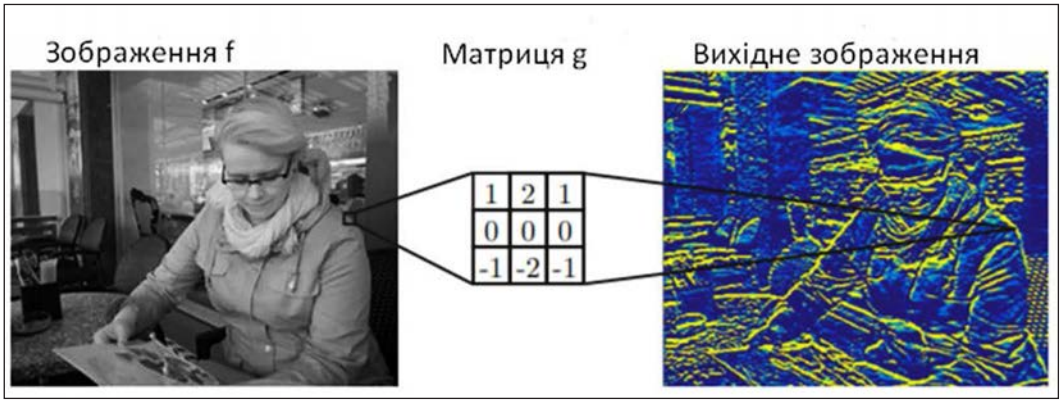

Рис. 1. Знаходження горизонтальних кутів на зображенні за допомогою згорткової нейронної мережі

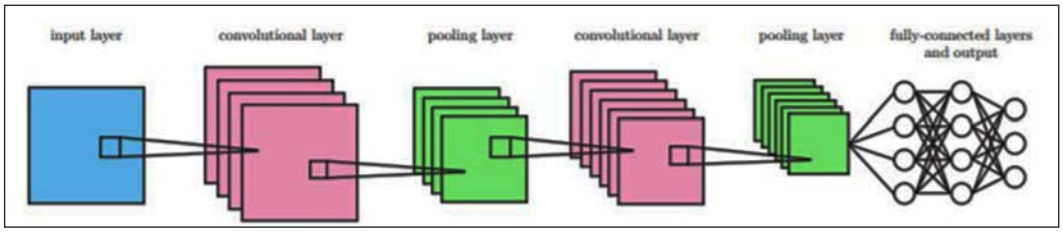

Рис. 2. Приклад згорткової нейронної мережі

глибокі шари мережі вимагають менше інформації про точні просторові місця розташування ознак, але потребують більше матриць фільтру, щоб розпізнати кілька шаблонів високого рівня. Завдяки зменшенню висоти та ширини об'єму даних ми можемо збільшити глибину обсягу даних і підтримувати час обчислень на розумному рівні. Є два способи зменшення обсягу даних. Один із способів – включити об'єднаний шар після згорткового шару. Шар ефективно знижує проби карти активації. Об'єднання має додатковий ефект, що робить отриману мережу більш інваріантною трансляцією, змушуючи детектори бути менш точними. Однак об'єднання може зруйнувати інформацію про просторові зв'язки між підрозділами моделей. Типовим методом об'єднання є максимальне об'єднання. Макс-об'єднання просто виводить максимальне значення в прямокутному районі карти активації.

Ще одним способом зменшення обсягу даних є коригування параметра кроку операції згортання. Параметр stride керує тим, чи обчислюється вихід згортки для мікрорайону, зосередженого на кожному пікселі вхідного зображення (stride 1), або для кожного n-го пікселя (stide n). Дослідження показали, що об'єднання шарів часто можна відкинути без втрати точності, використовуючи згорткові шари з більшим значенням кроку. Операція кроку еквівалентна використанню фіксованої сітки для об'єднання.

Важливим елементом цього підходу є так звані додаткові шари. Згортковий шар зазвичай включає нелінійну функцію активації, таку як випрямлена функція лінійної активації. Активації іноді описуються як окремий шар між згортковим шаром і шаром об'єднання.

Деякі системи також реалізують шар, який називається локальною нормалізацією реакції та використовується як метод регуляризації. Локальна нормалізація реакції імітує функцію біологічних нейронів, названу бічним гальмуванням, внаслідок чого збуджені нейрони знижують активність сусідніх нейронів. Однак інші методи регуляризації нині більш популярні, і вони обговорюються в наступному розділі.

Кінцеві приховані шари CNN, як правило, є повністю пов'язаними між собою. Однак для повного підключення шару потрібен досить невеликий обсяг даних. Налаштування об'єднання та кроку можна використовувати для зменшення розміру обсягу даних, який досягає повністю пов'язаних шарів. Звита мережа, яка не включає повністю пов'язаних шарів, називається повністю згортковою мережею (FCN). Якщо мережа використовується для класифікації, вона зазвичай включає софтмакс вихідний шар. Активації найвищих шарів можуть також використовуватися безпосередньо для створення функціональної репрезентації зображення. Це означає, що згорткова мережа використовується як великий детектор функцій.

Важливо також пояснити роль регуляризації. Вона стосується методів, що застосовуються для зменшення перевитрати, вводячи додаткові обмеження чи інформацію до системи машинного навчання. Класичний спосіб використання регуляризації в нейронних мережах – додавання штрафу до функції втрат, яка карає певні типи ваг. Особливість спільного використання параметрів згорткових мереж – ще один приклад регуляризації.

Висновки. Реалізація інтелектуальної системи, здатної розпізнавати конкретні об'єкти, – це вирішення складного завдання, що зачіпляє як технічний, так і математичний аспекти. Ця завдання вимагає поглибленого вивчення предметної галузі, аналізу проблеми, вміння і навичок у роботі з даними, глибоких знань у галузі дискретної математики, програмуванні, а також у багатьох інших сферах науки.

Використання згорткових нейронних мереж у розробленні інтелектуальної системи розпізнавання візуальних ключів дає змогу вирішувати проблеми перенасичення інформації, прискорити процес обробки даних і збільшити якісь та коректність аналізу.

Ця інтелектуальна система може використовуватися у сферах, пов'язаних з організацією великих подій та урбаністикою.

Список літератури:

1. Imagenet large scale visual recognition challenge 2016. URL: http://image-net.org/challenges/LSVRC/2016 (дата звернення: 2017.04.07).
2. Infotrends – how long does it take to shoot 1 trillion photos? URL: http://blog.infotrends.com/?p=21573 (дата звернення: 2019.06.17).
3. Matconvnet: Cnns for matlab. URL: http://www.vlfeat.org/matconvnet (дата звернення: 2019.01.10).
4. Software at the personal website of derek hoiem. URL: http://dhoiem.cs.illinois.edu/software (дата звернення: 2018.02.17).
5. Зайцев И.В. Нейронные сети: основные модели : учеб. пособие для физ. ф-та. Воронеж, 1991.
6. Bishop C.M. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
7. Головко В.А. Нейронные сети: обучения, организация и применение. Москва : ИПРЖР, 2008.

Oleksii A.V., Kornaha Ya.I. INTELLECTUAL SYSTEM OF RECOGNITION OF VISUAL KEYS FOR MOBILE APPLICATIONS

The world is constantly growing in the amount of image data and the rate of growth. In 2019, over 1.8 billion images are uploaded daily by top platforms such as Instagram and Facebook. As for video, according to Google, more than 86,000 hours of video are uploaded to YouTube every day. There are many different types of devices in the world that can take photos and videos. Personal smartphones and other devices are capable of capturing large-scale images, photo and video devices are available as never before. The video capturing equipment with automatic image capture is also quite affordable. The cars are equipped with cameras that allow you to effectively monitor traffic and prevent accidents. In order to effectively manage all this data, we need to have some idea of their content, namely the objects and details present in the image. Automated image content processing is useful for a wide variety of image-related tasks. For computer systems, this means crossing the so-called semantic divide between pixel information stored in a file and understanding the information content of an image. Automatic processing allows you to identify objects contained in image files, to process and to describe the details and parameters of these objects. This is called object detection and is one of the major problems with computer vision. Automatic systems and robots are critical to assessing the visual scene around you and being able to analyze the space around you, process information. As we demonstrate, nowadays, convolutional neural networks are a modern solution for object detection. This article will introduce the work of the intelligent visual key recognition system for mobile applications, namely the detection of objects on images using convolutional neural networks. It will also analyze the work of neural systems, identify how to teach this network to identify the objects required for the verification process, how to improve the operation of this algorithm through the self-learning neural network, the construction of the infrastructure of the system and what may be errors of the neural network in case of errors. In general, the article should address the issues of convolutional neural network design and its appropriateness.

Key words: *convolutional neural network, intelligent system, machine learning, algorithm, computer vision.*