КОМП'ЮТЕРНІ НАУКИ

УДК 519.6

ТЕХНІКИ НАВЧАННЯ НА МАЛОМУ НАБОРІ ДАНИХ ДЛЯ ЗАДАЧІ СЕГМЕНТАЦІЇ

М. Баранов, Ю. Щербина

Львівський національний університет імені Івана Франка,
вул. Університеська, 1, Львів, 79000,
e-mail: mapsg32@gmail.com, yshcherbyna@yahoo.com

За останні кілька років значного успіху було досягнуто у сфері глибокого навчання нейронних мереж для широкого спектра таких задач комп'ютерного бачення – класифікація, сегментації, локалізація тощо. Проте побудовані моделі досягають високих показників точності лише в певній обмеженій області даних, що описуються множиною тренувальних даних. Існує багато відкритих наборів даних (PASCAL VOC, ImageNet, COCO тощо) для різноманітних задач комп'ютерного бачення. Хоча ці набори даних охопили широкий спектр категорій об'єктів, існує все ще значна кількість об'єктів, які не входять в такі множини. Створення нового набору даних достатньо трудомісткий процес (особливо для задачі сегментації зображень) і, у підсумку, не завжди прийнятний у багатьох бізнес-моделях з огляду на фінансові затрати. Розглянуто технології побудови нейронних мереж для завдання сегментації зображень з використанням малої кількості проанотованих даних; побудовано архітектуру згорткової нейронної мережі на базі архітектури VGG-16 з використанням фреймворку Tensorflow 2 та натреновано на наборі даних FSS-1000. Головна ідея архітектури полягає у використанні множини підтримки як додаткових вхідних даних. Така множина містить декілька (у цій праці 5) зображень з відповідними масками сегментації. Натренована модель використовує цю множину як "приклади правильної сегментації". Набір даних було розділено на дві множини, що не перетинаються в сенсі категорій зображень. Запропонована архітектура, що тренувалась на 800 категоріях, досягає 83.55% точності на інших 200 категоріях зображень, що раніше не траплялись у процесі тренування моделі. Отримана модель здатна сегментувати довільні задані об'єкти на зображеннях. Запропонований підхід можна використовувати і для інших задач комп'ютерного бачення – класифікації, локалізації об'єктів і навіть генерації нових зображень.

Ключові слова: машинне навчання, нейронні мережі.

1. Вступ

Методи машинного навчання отримали широке застосування у різноманітних сферах сучасності завдяки гнучкості алгоритмів і можливості адаптації до довільних даних.

Як механізм узагальненої обробки даних, широкого використання набули нейронні мережі. Значного прогресу було досягнуто у різноманітних галузях комп'ютерного бачення (задачі класифікації, сегментації, генерації, анотації зображень, розпізнавання об'єктів, тексту, жестів рук тощо), обробки природної мови (розуміння мови, генерація мови, конвертація голосу в цифровий текст, озвучення тексту тощо) та в інших галузях машинного навчання. Глибокі нейронні мережі потребують масштабних наборів даних для уникнення ефекту перенавчання. Незважаючи на велику кількість відкритих даних, доступних для використання, значна кількість задач залишається нерозв'язною через брак даних.

Збір та анотація нового набору даних для унікальної задачі потребує значної кількості людської праці. Особливо багато часу на анотацію одного зображення витрачається для створення маски сегментації. В багатьох випадках виготовлення нового набору даних економічно невигідно. З іншого боку, існують такі задачі, для яких збір масштабних наборів даних неможливий (наприклад, задача передбачення динаміки розвитку живих організмів в умовах невагомості тощо). Варто також відмітити низку медичних задач: в більшості країн дані пацієнтів конфіденційні та не підлягають оприлюдненню, що значно ускладнює розробку моделей з використанням глибоких нейронних мереж у сфері медицини. Варто зауважити, що задача сегментації зображення дуже популярна для медичних задач (сегментація ураженої області тканини тощо).

З іншого боку, більшість бізнес моделей розглядають постійну адаптацію моделей машинного навчання до нових даних (наприклад, розпізнавання продуктів у супермаркеті). Перетренування класичних моделей може займати значний час (до декількох тижнів машинного часу), не говорячи вже про неможливість оновлення наборів даних за короткий час.

Обмеження у розмірі наборів тренувальних даних і значна кількість часу на тренування моделей глибокого навчання призводить до необхідності побудови такої архітектури нейронної мережі, яка інваріантна до вхідних даних і може працювати з довільними, наперед не заданими об'єктами.

2. Задача сегментації зображень

В області комп'ютерного бачення сегментація зображень – це процес поділу цифрового зображення на множину сегментів (множини пікселів, які також називають об'єктами на зображенні). Ціль сегментації – спростити і змінити семантичну репрезентацію зображення на таку, яка несе більше інформації і може бути легше опрацьована в майбутньому. Сегментація об'єктів зазвичай використовується для пошуку та локалізації різноманітних об'єктів на зображенні та визначення їхніх контурів (ліній, кривих тощо). Більш точно, сегментація зображення – це процес призначення міток кожному пікселю зображення так, що пікселі з однаковою міткою мають конкретні спільні характеристики. Розрізняють два типи сегментації зображення:

– семантична сегментація (англ. semantic segmentation);
– сегментація екземплярів (англ. instance segmentation).

Результатом семантичної сегментації загалом є мітка класу для кожного пікселя, тоді, як сегментація екземплярів дає змогу з'ясувати відношення пікселів до того чи іншого об'єкта без розуміння категорії цього об'єкта. Іншими словами, семантична сегментація не може відділити два автомобілі, які стоять поруч. Сегментація екземпляра успішно відрізняє ці два об'єкти, без розуміння, що це автомобілі. На практиці зазвичай використовують моделі, які здатні будувати одночасно і семантичну сегментацію, і сегментацію екземплярів.

Існує багато алгоритмів для сегментації зображень. Найпростіші алгоритми ґрунтуються на аналізі яскравості зображення – порогові операції. Очевидно, що кожна конкретна задача потребує унікального алгоритму. До того ж такі алгоритми чутливі до зовнішніх чинників (наприклад, яскравість зображення тощо).

З найпростіших статистичних методів часто використовують K-Means кластеризацію, трактуючи кожен піксель як $3D$ точку (три складові кольору в кольоровому

просторі RGB, HSV тощо). Такий алгоритм значно загальніший, проте не є придатним для сегментації складних об'єктів – людина, автомобіль тощо.

Для складних задач сегментації зазвичай використовують згорткові нейронні мережі.

3. Класичні підходи машинного навчання для задачі сегментації

Машинне навчання широко застосовується у сфері комп'ютерного бачення для низки задач – класифікація, сегментація зображень, пошук відомих об'єктів, побудова карт глибин тощо. Сучасні зображення містять велику кількість інформації. Наприклад, класичний набір даних MNIST [7] містить чорно-білі зображення розміром 28×28 пікселів. А це 784 байти даних лише на один приклад. Сучасні ж зображення зазвичай мають розширення не менше FullHD (1920×1080). Враховуючи таку колосальну кількість інформації для обробки, необхідно використовувати глибокі нейронні мережі (англ. deep learning). Завдяки детальному дослідженню питання застосування методів градієнтного спуску розроблено алгоритми [25,31], які успішно застосовані для тренування глибоких нейронних мереж. Також згорткові нейронні мережі демонструють великий потенціал у сфері комп'ютерного бачення завдяки концепції операції згортки в обробці просторових сигналів [3–5, 30]. Отож не дивно, що згорткові нейронні мережі широко застосовують для задачі сегментації зображень.

4. Проблема масштабних наборів даних

Наявність масштабних наборів даних дає змогу експерементувати з глибокими нейронними мережами. Незважаючи на значну кількість категорій проанотованих у таких наборах даних як PASCAL VOC [8] (19 740 зображень, 20 класів), ILSVRC [26] (1 281 167 зображень, 1000 класів), COCO [17] (204 721 зображень, 80 класів) – значна частка категорій об'єктів залишається непроанотованою. Груба оцінка кількості різноманітних об'єктів на землі потрапляє в межі від 500 000 до 700 000 класів (відповідно до загальної кількості іменників англійської мови). Наведені вище великомасштабні набори даних не покривають і 1% загальної кількості усіх об'єктів. Додавання нового класу до існуючого набору даних потребує великої кількості людської праці (особливо для задачі сегментації, де кожен об'єкт треба покрити маскою довільної форми). Наприклад, середня кількість зображень на один клас у наборі даних ImageNet – 650. Варто зауважити, що на одному зображенні може міститись від одного до декількох сотень (і навіть тисяч) різноманітних об'єктів.

З іншого боку, окрім необхідності анотацції великих об'єктів даних, прикладне застосування нейронних мереж у бізнес моделях має й іншу проблему – перетреновування моделі. Кожен раз при додаванні нового класу в загальному випадку є необхідність перетреновувати модель з нуля. Звичайно, існують техніки, які спрощують процес перетренування, але в будь-якому випадку це потребує затрати часу та обчислювальних ресурсів.

5. Техніки навчання на малому наборі даних

Техніки навчання на малому наборі даних (англ. few shot learning) широко відомі для задач класифікації зображення. Загалом такі техніки поділяють на три типи:

– на основі моделі;
– на основі алгоритмів оптимізації;
– на основі метрик.

Широко відома техніка тонкого налаштування полягає в тренуванні уже наперед натренованої мережі. Припустимо, що потрібно натренувати мережу для розпізнавання моделей автомобілей. Можна використати існуючий великомасштабний набір даних (наприклад, ImageNet). Опісля, використовуючи ваги натренованої моделі як початкові ваги, отримують кращий результат тренування і досягнуто його буде у коротший час. Такий процес називається "передача знань" (англ. transfer learning). Якщо ж зафіксувати усі ваги мережі і тренувати лише декілька останніх шарів, то результат буде досягнуто ще швидше. Цей прийом називають "точним налаштуванням" (англ. fine-tuning).

Детальне дослідження "точного налаштування" моделей проведено у [29], де доведено, що використання косинусної відстані та обережного підбору гіперпараметрів дає змогу отримувати хороші результати. З використанням моделі Faster RCNN [23] на PASCAL VOC досягнуто 57% mAP при 10 прикладах на клас.

MAML [10] (англ. model agnostic meta learning) пропонує використання другої похідної для тренування моделі для розпізнавання багатьох класів об'єктів одночасно (одночасно декілька класів за один крок оптимізації). Отож, ваги отриманої моделі легко адаптуються на суміжні завдання (класи об'єктів). Така техніка може бути застосована до будь-якої моделі та задачі, якщо модель тренується методами градієнтного спуску. Необхідність використання другої похідної унеможливлює використання такої техніки для великих моделей. FOMAML [12] алгоритм розширює MAML алгоритм, використовуючи лише першу похідну. Так FOMAML алгоритм придатний для застосування в задачі сегментації зображень.

Техніки навчання на малому наборі даних на основі метрик (англ. metric learning) найбільш поширений спосіб. Існує багато підходів до класифікації зображень, які використовують дескриптори. Як з'ясовано у [14, 19], в процесі тренування беруть участь одночасно три приклади зображень. Один випадковий приклад з набору даних A – якір (англ. anchor), довільне зображення того ж класу P – позитивний приклад (англ. positive) і довільний приклад іншого класу N – негативний приклад (англ. negative).

$$A, P, N \in D,$$

де D – набір даних.

Нехай модель M продукує дескриптор зображення d, $(M(X) = X_d$. У процесі тренування мінімізується функція втрат

$$max(||M(A) - M(P)||_L - ||M(A) - M(N)||_L + \alpha, 0).$$

Так збільшується відстань між дескрипторами зображень різних класів і мінімізується відстань дескрипторів зображень одного класу. Сформована база даних дескрипторів дає змогу класифікувати зображення з використанням лише його дескриптора. Важливим є те, що база дескрипторів може поповнюватись незалежно від моделі.

Як дескриптор можна використовувати різні функції. Зазвичай використовують MAC [21], глобальне стягування (усереднене, максимальне). Потужний потенціал демонструє дескриптор узагальненого параметричного стягування GeM [19].

Використання змішаних дескрипторів поліпшує результати попередніх досягнень. Комбінацію GeM, стягування, MAC опрацьовано в [11].

6. Техніки навчання на малому наборі даних для завдання пошуку та сегментації

Задача пошуку та сегментації зображень значно складніша, оскільки, окрім класу, треба з'ясувати просторове положення об'єкта. Світовий досвід технік навчання на малому наборі даних для задачі пошуку об'єктів та сегментації зображень на порядок менший порівняно з задачею класифікації. Техніки, які застосовують до задач класифікації, не можна безпосередньо застосовувати до задач сегментації, проте ідеї цих технік успішно переносяться на задачі сегментації.

Для задачі сегментації широко використовують дві множини вхідних даних – множина запиту (зображення для сегментації) та множина підтримки (множина проанотованих зображень, приклади правильної сегментації). Наприклад, модель YOLOv2 [22] адаптована для техніки навчання на малому наборі даних завдяки модулю переважування [6]. Множина підтримки генерує вагові коефіцієнти, які в режимі реальної роботи (не тренування) модифікують ваги мережі для адаптації до нового класу об'єктів. Отже, змінюючи лише множину підтримки (до десяти зображень з відповідними анотаціями), можна адаптувати модель до розпізнавання об'єктів нових категорій.

Ідея сіамських нейронних мереж [20] може бути перенесена на модель Faster RCNN/RCCN Masker [23]. Після модуля витягування ознак дескриптори зображень (які охоплюють дані про просторове положення об'єктів множини підтримки) змішуються і решта моделі, оперуючи цими даними, генерує відповідну маску об'єктів для зображення з множини запиту.

Рис. 1. Набір даних FSS-1000

7. Набір даних fss-1000

FSS-1000 [32] – набір даних для тренування моделей для задачі сегментації. Дані містять 1000 різноманітних категорій повсякденних об'єктів. Класи об'єктів між собою абсолютно незалежні. Дані збирали з використанням різних пошукових систем (Google, Bing, Yahoo). Основна ідея цього набору даних – немає прив'язки до конкретних об'єктів і рівномірна різноманітність даних. Наприклад, у наборі даних

OpenImages є декілька десятків категорій порід собак, що безумовно відображається в результуючі моделі – покращене вміння розпізнавати собак і слабкість до інших об'єктів. Другий акцент FSS-1000 – рівномірний покласовий розподіл кількості даних – на кожен клас припадає точно 10 зображень з відповідними масками сегментацій. На одному зображені міститься лише одна категорія об'єктів. Анотації містять лише маску об'єктів без розділення на окремі екземпляри. Усі зображення стиснені до розміру 224×224 пікселів. Ті зображення, які в оригіналі мали співвідношення сторін менше 0,5 або більше 2, відкинули. Такий набір даних дуже хороший для тренування моделей з використанням техніки навчання на малому наборі даних. На рис. 1 зображено приклади зображень з набору даних FSS-1000.

8. Запропонована архітектура мережі

Модель складається з трьох основних блоків [2, 32]. Блокова архітектура моделі зображена на рис. 1.

Модуль витягування ознак. Модуль витягування ознак відповідає за генерування карти ознак вхідного зображення. Карта ознак використовується для генерації результуючої маски сегментації.

Модуль відношення. Цей модуль важливий в техніці навчання на малому наборі даних. Тут відбувається об'єднання репрезентацій зображень двох множин – запиту та підтримки.

Модуль генерування маски сегментації. Цей модуль на вході отримує об'єднані репрезентації множини запиту і множини підтримки і на виході видає результат.

На рис. 2 зображено схему архітектури моделі. Архітектура реалізована з використанням фреймворку TensorFlow 2 [1].

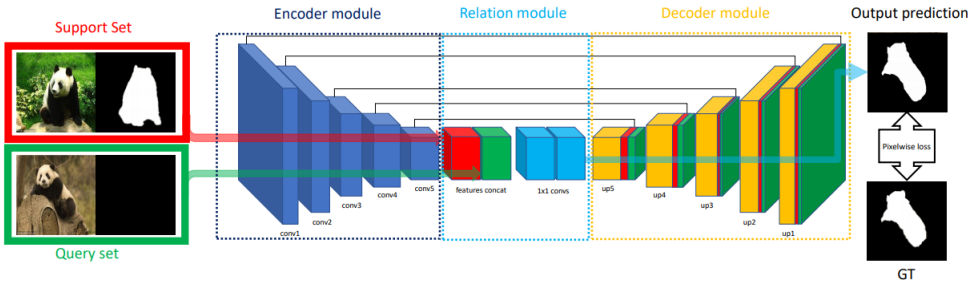

Рис. 2. Архітектура моделі. Модель складається з кодувальника (encoder), декодувальника (decoder) та модуля відношення (relation module). На вхід модель отримає множину запиту (query) та підтримки (support)

За основу модуля витягування ознак було обрану архітектуру VGG-16 запропоновану К. Сімонаном і А. Цісерманом [28]. VGG-16 – згорткова нейронна мережа, яка поліпшує архітектуру AlexNet [16], заміняючи великі ядра згорткових шарів на менші. Архітектура моделі складається зі згорткових блоків – групи трьох згорткових шарів. Детальний опис мережі наведено у табл. 1. Початкові ваги були скопійовані з натренованої мережі VGG-16 на наборі даних ImageNet [26], оскільки для випадкових побутових об'єктів більшість фільтрів можна успішно перевикористати.

Таблиця 1

Архітектура VGG-16

#	**Тип шару**	**Ядро**	**Фільтри**
1	Згортковий	3×3	64
2	Згортковий	3×3	64
3	Стягування	2×2	64
4	Згортковий	3×3	128
5	Згортковий	3×3	128
6	Стягування	3×3	128
7	Згортковий	3×3	256
8	Згортковий	3×3	256
9	Згортковий	3×3	256
10	Стягування	2×2	256
11	Згортковий	3×3	512
12	Згортковий	3×3	512
13	Згортковий	3×3	512
14	Стягування	2×2	512
15	Згортковий	3×3	512
16	Згортковий	3×3	512
17	Згортковий	3×3	512
18	Стягування	2×2	512

Ефективна репрезентація зображення буде використовуватись як основа для генерації маски сегментації. Модуль відношення має поєднувати карти ознак множини запиту та підтримки. На практиці реалізація такого модуля найпростіша. Нагадаємо, що ефективна репрезентація зображень – це множина матриць (тензор) розміру $W \cdot H \cdot C$, де C – кількість каналів. Об'єднання двох репрезентацій відбувається шляхом конкатенації двох тензорів по глибині C.

Архітектура цього модуля реалізована на основі архітектури VGG-16 в зворотному порядку: починаючи від останнього шару до першого. Шари максимального стягування замінені на операцію збільшення матриць з використанням інтерполяції. Вхідними даними модуля генерації є конкатеновані карти ознак з модуля відношення. В процесі тренування саме цей блок моделі виявляє відповідності між двома картами ознак множини запиту та підтримки.

Нагадаємо, що двовимірна згортка з ядром $m \cdot n$ насправді використовує ядро згортки на одну розмірність більше: $m \cdot n \cdot K$, де K – вхідна кількість фільтрів карти ознак. Враховуючи той факт, що модуль відношення конкатенує карту ознак по глибині, подальші згортки утворюють лінійні комбінації карти ознак множини запиту та підтримки, установлюючи відповідність. Карта ознак множини запиту містить інформацію про правильні маски сегментацій прикладів об'єктів. Саме цей факт і допомагає моделі успішно генерувати маски сегментації для множини підтримки.

В архітектурах згорткових мереж зазвичай використовують шари стягування (наприклад, максимального стягування). Такі шари значної мірою збільшують рецептивне поле мережі, але, з іншого боку, зменшують чутливість до деталей. Така

характеристика не дає змогу модулю генерації згенеруати точну маску сегментації. Проте ми точно знаємо, що карта ознак містила найдрібніші деталі на перших згорткових шарах. Якщо передати ці карти з модуля витягування карти ознак в модуль генерації масок, то так можна значно збільшити точність моделі. Проте передавати ці дані треба особливо між відповідними шарами. Оскільки модуль генерації створений відповідно до оберненого модуля генерації ознак, то репрезентації перших шарів потрібно надіслати останнім шарам модуля генерації масок. Відповідно, ознаки з середини модуля – до середини блока генерації тощо.

Загалом в архітектурі використано п'ять таких з'єднань швидкого доступу. Ідея з'єднань швидкого доступу отримана з архітектури ResNet [15] (Residual block) та U-Net [24].

9. Набір даних для тренування

Для тренування моделі використано набір даних FSS-1000 [32]. Дані було розбито на дві категорії – тренувальні та тестувальні дані в пропорції 80% тренувальної вибірки та 20% тестувальної вибірки. Розбиття відбувалось так, щоб категорії, які використовують в процесі тренування, не використовувались на етапі тестування. Загалом у тренувальній вибірці виявилось 800 категорій, а в тестувальній – 200. Кожну категорію для обох вибірок було розділено на дві множини – множину запиту та множину підтримки. Кількість зображень у множині запиту – 5; 5 інших зображень використували для тренування (або тестування, залежно від вибірки).

З додаткових даних використовувалась вибірка ImageNet [26] для отримання натренованих ваг модуля витягування ознак (VGG-16). Усі інші шари нейронної мережі було ініціалізовано випадковими початковими вагами.

10. Результати та метрики

У процесі тренування ми оптимізуємо функцію втрат. Значення функції втрат не є абсолютним показником і не може бути використано для визначення якості натренованої моделі. Для цього використовують інші метрики – точність, повнота, F1-Score, IoU тощо.

Уведемо такі позначення:

- tp – кількість правильно розпізнаних пікселів на множині пікселів об'єкта;
- fp – кількість неправильно розпізнаних пікселів на множині об'єкта;
- tn – кількість правильних пікселів на множині фону зображення;
- fn – кількість хибно розпізнаних пікселів на множині фону.

Використовуючи описані вище позначення, визначимо такі метрики:

1. Точність

   $Precision = \frac{tp}{tp+fp}$.

2. Повнота

   $Recall = \frac{tp}{tp+fn}$.

3. F1-Score

   $F1Score = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$.

4. IoU

   $IoU = \frac{tp}{tp+fp+fn}$.

Таблиця 2

Результати експериментів

Метрика	**Значення**
Точність	**83,55%**
Повнота	**81,49%**
F1-Score	**82,51%**
IoU	**68,73%**

Таблиця 3

Порівняння метрик на тестових категоріях набору даних FSS-1000

Архітектура	**IoU**	**Час обробки**
Adaptive relation network	**80,12%**	0.0338c
Запропонована архітектура	68.73%	**0.0234c**

Усі виміри було проведено на тестувальній вибірці, тобто на категоріях об'єктів, які є невідомими моделями (тобто не використовувались для тренування). Розмір множини підтримки 5. Результати отримані в процесі експериментів описані в табл. 2. Порівняння запропонованої архітектури з базовим підходом, описаним у [32], наведено у табл. 3.

Натренована модель здатна сегментувати довільні об'єкти, задані у множині підтримки. Отож, для додавання нового класу для сегментації достатньо отримати анотації лише 5 зображень нового об'єкта. Більше того, таке додавання нової категорії об'єктів не потребує перетренування моделі. Тобто, додавання нової категорії потребуватиме лише кількох хвилин людського часу на відмінну від кількох місяців анотування великомасштабних наборів даних.

Наведена архітектура моделі у цій праці містить два обмеження:

1. Семантична сегментація.

   Модель здатна прокласифікувати попіксельно зображення, але неможливо точно відрізнити два окремих екземпляри (особливо, коли вони дотикаються один до одного).

2. Сегментація лише одного класу за один прохід.

   Оскільки модель сегментує об'єкт, зазначений у множині підтримки, то за один прохід моделі можемо отримати маску лише для одного класу. Для сегментації іншого класу треба проводити ще один запуск моделі.

Для генерацій карти ознак перший модуль моделі використовується однакового для множини запиту та підтримки, тобто є спільним. Елементи множини підтримки містять маску анотації, а отже, усі елементи – чотириканальні зображення. Це створює необхідність замінити перший шар VGG-16 [28] на новий шар (не натренований), який приймає чотириканальні зображення. Для загальності множина запиту також є множиною чотириканальних елементів. Оскільки маска об'єктів для елементів запиту невідома, то четвертий канал заповнюється нулями. Множина запиту може містити довільну кількість зображень. Для успішної конкатенації

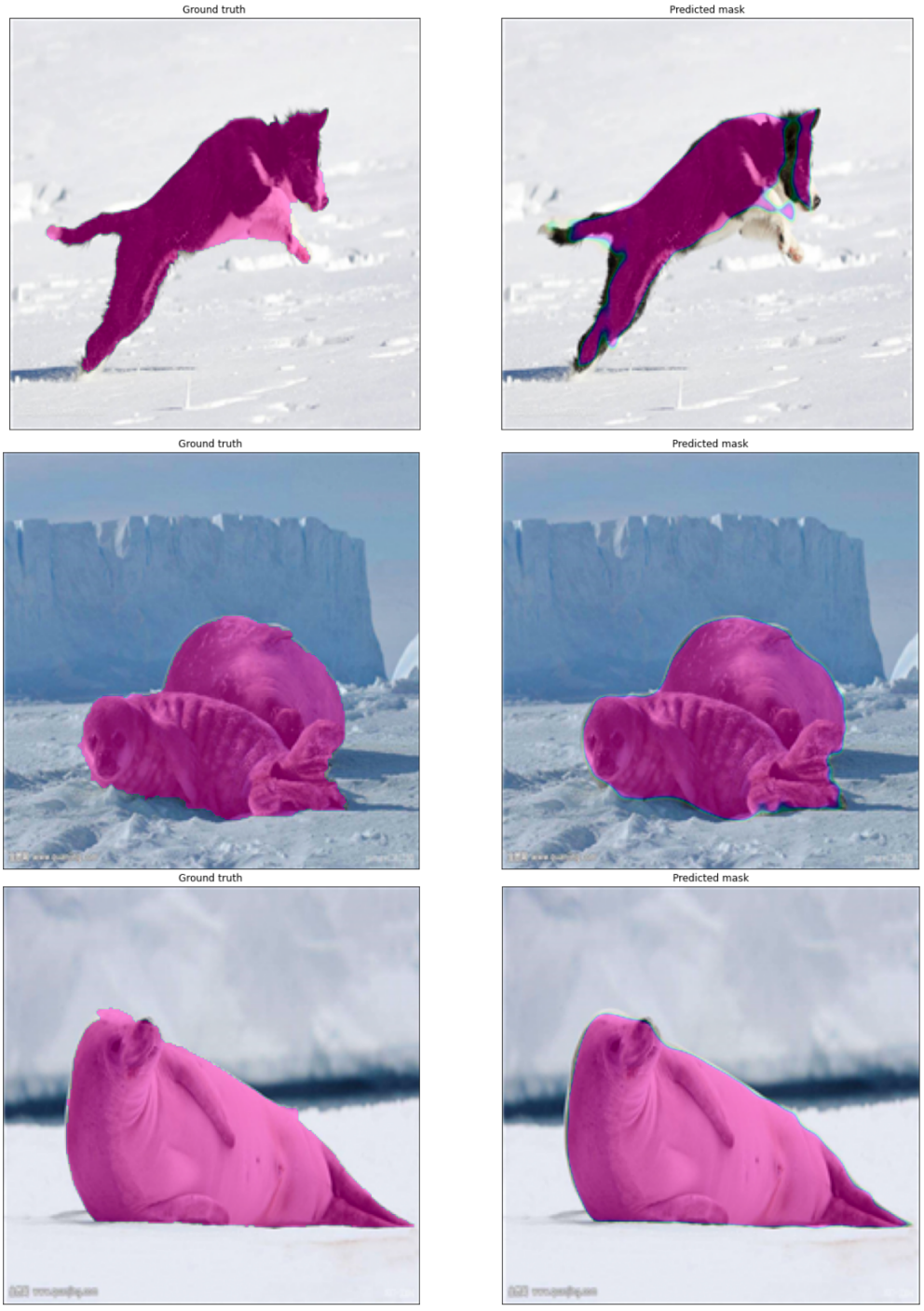

Рис. 3. Приклади сегментації

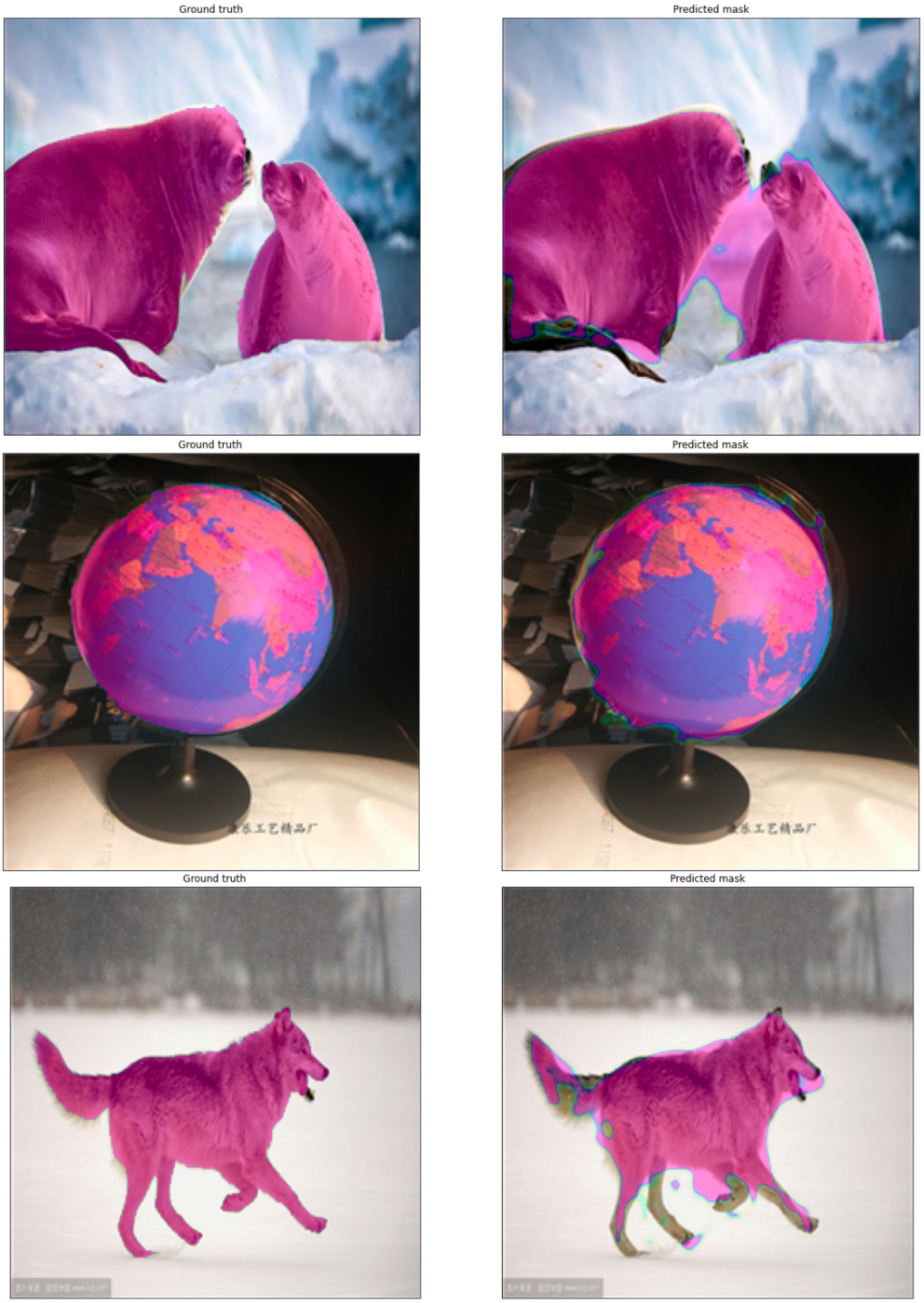

Рис. 4. Приклади сегментації

карти ознак і отримування входу сталого розміру на останній блок моделі усі карти ознак для кожного прикладу з множини підтримки підсумовуються поелементно. Множина запиту може містити лише одне зображення за один прохід.

11. Висновки

У цій праці досліджено потенціал згорткових нейронних мереж, здатних виконувати тренування на малій кількості даних. Основні результати роботи такі.

1. Досліджено математичний апарат та ідею навчання на малому наборі даних. Розкрито важливість використання розбиття вхідних даних на множину запиту та підтримки. Досліджено вплив множини підтримки на результат тренування та роботи моделі.
2. Запропоновано архітектуру згорткової нейронної мережі з використанням множин запиту та підтримки. Реалізована архітектура об'єднує ідеї таких моделей: VGG-16 [28] та U-Net [24]. Важливим внеском є адаптація VGG-16 [28] для множини підтримки й уніфікація для множини підтримки.
3. Досліджено та доведено на практиці можливість адаптації розглянутої архітектури для сегментації зображень категорій об'єктів, які не були відомими на етапі тренування.

Список використаної літератури

1. *Баранов М.* Бібліотека Tensorflow для задач машинного навчання / М. Баранов // Дев'ята науково-практична конференція FOSS Lviv. – Збірник наукових праць. – Львів, 2019. – С. 7–9.
2. *Баранов М.* Техніка навчання на малому наборі даних для задачі сегментації / М. Баранов // Міжнародна студентська наукова конференція з прикладної математики та комп'ютерних наук МСНКПМКМ-2020. – Львів, 2020. – С. 24–29.
3. *Баранов М.* Нейронні мережі для розпізнавання рукописного тексту / М.Баранов // Сучасні проблеми прикладної математики та інформатики. – XXV міжнародна наукова конференція. – Львів, 2019. – С. 8–12.
4. *Баранов М.* Нейронні мережі для розпізнавання рукописного тексту / М. Баранов // Міжнародна студентська наукова конференція з прикладної математики та комп'ютерних наук СНКПМКМ-2019. – Львів, 2019. – С. 61–64.
5. *Нікольський Ю.* Системи штучного інтелекту / Ю. Нікольський, В. Пасічник, Ю. Щербина // Навчальний посібник. Серія "Комп'ютинг". – Львів: Магнолія, 2006. – 2013. – 278 с.
6. *Bingyi K.* Few-shot object detection via feature reweighting / K. Bingyi, L. Zhuang, W. Xin., Y. Fisher, F. Jiashi, D. Trevor // Proceedings of the IEEE International Conference on Computer Vision. – 2019. – С. 8420–8429.
7. *Feiyang Chen* Assessing four neural networks on handwritten digit recognition dataset (MNIST) / Chen Feiyang, Chen Nan, Mao Hanyang, Hu Hanlin // arXiv preprint arXiv: 1811.08278. – 2018.
8. *Everingham M.* The Pascal Visual Object Classes (VOC) Challenge / M. Gool Luc Van Everingham, Christopher K. I. Williams, J. Zisserman A. Winn // International Journal of Computer Vision. – Springer Verlag. – 2009. – P. 303–308.
9. *Fabricio B.* Interactive image segmentation using label propagation through complex networks / B. Fabricio // Expert Systems With Applications. – Elsevier. – 2019. – P. 18–33.
10. *Finn Chelsea* Model-agnostic meta-learning for fast adaptation of deep networks / Chelsea Finn, Pieter Abbeel, Sergey Levine // arXiv preprint arXiv: 1703.03400. – 2017.

11. *HeeJae Jun* Combination of Multiple Global Descriptors for Image Retrieval / Jun HeeJae, Ko ByungSoo, Kim Youngjoon, Kim Insik, Kim Jongtack // arXiv. – 2019.
12. *Hendryx Sean M.* Meta-learning initializations for image segmentation / Sean M. Hendryx, Andrew B. Leach, Paul D. Hein, Clayton T. Morrison // arXiv preprint arXiv: 1912.06290. – 2019.
13. *Jiawei J.* K-Means Clustering / J. Jiawei, X. Jiawei, H. Jiawei // Springer US. – Boston, MA. – 2010. – P. 563–564.
14. *Hermans Alexander* In defense of the triplet loss for person re-identification / Alexander Hermans, Lucas Beyer, Bastian Leibe // arXiv preprint arXiv: 1703.07737. – 2017.
15. *Kaiming H.* Deep residual learning for image recognition / H. Kaiming, Z. Xiangyu, R. Shaoqing, S. Jian // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 770–778.
16. *Krizhevsky A.* ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, E. Hinton Geoffrey // NIPS. – 2012.
17. *Lin T.* Microsoft COCO: Common objects in context. arXiv 2014 / T. Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev., Girshick, B. Ross, James Hays, Pietro Perona, Deva Ramanan, Dollár Piotr Zitnick C. Lawrence // arXiv preprint arXiv: 1405.0312. – 2014.
18. *Michaelis C.* One-Shot Instance Segmentation / I. Ustyuzhaninov, M. Bethge, A. Ecker // arXiv. – 2018.
19. *Radenović Filip* Fine-tuning CNN image retrieval with no human annotation / Filip Radenović, Giorgos Tolias, Chum Ondřej // IEEE transactions on pattern analysis and machine intelligence – видавництво IEEE, 2018. – P. 1655–1688.
20. *Ramachandra Bharathkumar* Learning a distance function with a Siamese network to localize anomalies in videos / Ramachandra Bharathkumar, Jones Michael, Vatsavai Ranga // The IEEE Winter Conference on Applications of Computer Vision. – 2020. – P. 2598–2607.
21. *Razavian Ali S.* Visual instance retrieval with deep convolutional networks / Ali S. Razavian, Josephine Sullivan, Stefan Carlsson, Atsuto Maki // ITE Transactions on Media Technology and Applications. – видавництво The Institute of Image Information and Television Engineers. – 2016. – P. 251–258.
22. *Redmon J.* YOLO9000: Better, faster, stronger. arXiv 2016 / J. Redmon, A. Farhadi // arXiv preprint arXiv: 1612.08242. – 2016.
23. *Shaoqing Ren* Faster r-cnn: Towards real-time object detection with region proposal networks / Ren Shaoqing, He Kaiming, Girshick Ross, Sun Jian // Advances in neural information processing systems. – 2015. – P. 91–99.
24. *Ronneberger O.* U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // International Conference on Medical image computing and computer-assisted intervention. – Springer, 2015. – P. 234–241.
25. *Rumelhart D. E.* Learning representations by back-propagating errors / D. E. Rumelhart, G. E. Hinton, R. J. Williams // Journal Nature. – 1983. – P. 533–536.
26. *Russakovsky O.* Imagenet large scale visual recognition challenge / O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein // International Journal of Computer Vision. – Springer, 2015. – P. 211-252
27. *Samuel A. L.* Some Studies in Machine Learning Using the Game of Checkers / A. L. Samuel // IBM Journal of Research and Development. – 1959. – P. 210-229.
28. *Simonyan K.* Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // arXiv preprint arXiv: 1409.1556. – 2014.
29. *Wang Xin* Frustratingly Simple Few-Shot Object Detection / Xin Wang, Thomas E. Huang., Trevor Darrell, Gonzalez, E. Joseph, Yu Fisher // arXiv preprint arXiv: 2003.06957.
30. *Wang Zijie J.* CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization / Zijie J. Wang., Robert Turko, Omar Shaikh, Haekyu Park, Nilaksh Das, Fred Hohman, Minsuk Kahng, Chau Duen Horng // arXiv preprint arXiv: 2004.15004. – 2020.

31. *Xavier G.* Understanding the difficulty of training deep feedforward neural networks / G.Xavier, B. Yoshua // Proceedings of the thirteenth international conference on artificial intelligence and statistics. – 2010. – P. 249–256.
32. *Xiang Li* FSS-1000: A 1000-Class Dataset for Few-Shot Segmentation / Li Xiang, Wei Tianhan, Yau Pun Chen, Yu-Wing Tai, Chi-Keung Tang // CVPR. – 2020.

Стаття: надійшла до редколегії 26.08.2020
доопрацьована 14.09.2020
прийнята до друку 23.09.2020

FEW-SHOT LEARNING FOR IMAGE SEGMENTATION TASK

M. Baranov, Yu. Shcherbyna

Ivan Franko National University of Lviv,
Universytetska str., 1, Lviv, 79000, Ukraine,
e-mail: mapsg32@gmail.com, yshcherbyna@yahoo.com

Over the past few years, we have witnessed the success of deep learning in various computer vision tasks such as image classification, object detection, object localization, etc. The existing approach achieves very good results only in a specific domain due to the dataset variability limitation. Today we have a lot of free open-source datasets (PASCAL VOC, ImageNet, COCO тощо) suitable for a variety of computer vision tasks. Although these datasets have covered a wide range of object categories, there is still a significant number of objects that are not included. Creating a new dataset is a very time-consuming process and may be not available for many business cases due to the huge cost of work. In this work, several few-shot image segmentation approaches were investigated. We have built a convolutional neural network based on the VGG-16 model (pretrained on the public available ImageNet dataset) using Tensorflow 2 and have trained it on the public available FSS-1000 dataset. The key idea of our approach is to use the support set as an additional input to the model. These sets can be retrieved as example cases for the model because the support set contains annotated images. Our model peeks to the support set and use information in that set as an example. That helps the model to pay attention only to features specific to the concrete object category. Model architecture consists of thee main blocks - encoder, relation module and the decoder module. Encoder process both input image and the support set and provide features to the relation module. Relation module combine both features from the input image and support set and then pass in to encoder module that produces final segmentation mask. Skip-connections is used as it proposed by the U-Net network. The FSS-1000 dataset was splitted into the train and test set with respect the image category. So, 2000 test images have no intersection with the rest of the 800 training categories. Our trained model is able to segment an arbitrary image using only a few annotated images passed to the network in the support set. We achieve up to 83,55% accuracy using novel object categories. Least but not last out approach can be adapted to the other computer vision tasks such as image classification, object detection and even image generation.

Machine learning – is a subset of artificial intelligence algorithms. The main purpose of such methods is not a solution of certain problem but building an algorithm to solve that problem. The main idea of machine leaning is based on statistical models. Term “machine learning” was introduced by Samuel Arthur in 1959 [27]. Machine learning is an important part of modern business and research. There are many scopes of application of the statistical models and various artificial intelligence methods. Machine learning approaches are divided into the following subsets:

- Supervised learning
- Unsupervised learning

Unsupervised learning considers such models which can be trained by itself only consuming data provided by human (without annotation). Despite the best advantages of such models supervised models are more popular nowadays. For example, in classification tasks supervised learning allows models that maps sample to certain class; unsupervised models can only distinct different group of samples without any knowledge about which class those samples belong to (such process is also called clusterization).

Neural networks become the most popular algorithm. Great success was achieved in different subsets of artificial intelligence using neural networks in computer vision (image classification, segmentation generation, annotation tasks, object detection, gesture recognition etc). Also, there is good progress in another fields.

In this work we present neural network architecture for image segmentation that can be trained using only few samples per class. That model consumes query image (image to be segmented) along with support set (examples of segmented images). This architecture is inspired by U-Net [24]. VGG-16 [28] is used as a network backbone. We trained our model using FSS-1000 dataset [32] and achieved about 69% IoU.

Key words: machine learning, Neural networks.