УДК 004.93

М. В. Бездітний
Р. В. Маслій

СЕМАНТИЧНА СЕГМЕНТАЦІЯ ЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Вінницький національний технічний університет;

Анотація

В статті проведено дослідження семантичної сегментації зображень з використанням згорткових нейронних мереж та середовища DIGITS. Для досліджень використані бази зображень PASCAL VOC та NVIDIA-Aerial Drone Dataset. В результаті навчання було досягнуто вірогідності сегментації у 83% та 99% відповідно.

Ключові слова: семантична сегментація, згорткова нейронна мережа, DIGITS, PASCAL VOC.

Abstract

In this article investigates the semantic segmentation of images using spinel neural networks and the DIGITS environment. The research uses the PASCAL VOC and NVIDIA-Aerial Drone Dataset images databases. As a result of the training, the probability of segmentation was achieved in 83% and 99% respectively..

Keywords: semantic segmentation, convolutional neural network, DIGITS, PASCAL VOC.

Вступ

В останні декілька років нейронні мережі отримали широке розповсюдження. Нейронні мережі можуть вирішувати досить складні та різноманітні задачі, такі як: розпізнавання та класифікація об’єктів у зображеннях, розпізнавання тексту, розпізнавання голосу, виявлення неполадок у складних системах, оцінка ризиків у банківській сфері тощо.

Зростання апаратних можливостей комп’ютерів з залученням можливостей GPU і наявність бібліотек та середовищ для навчання нейронних мереж уможливлюють дослідження нейронних мереж на персональному комп’ютері.

Задача семантичної сегментації зображень полягає у виділенні у зображеннях локальних областей, які відповідають різним класам об’єктів. Ця задача є більш складнішою чим задача класифікації зображень чи пошуку об’єктів у зображеннях, так як це обумовлено не тільки необхідністю визначення класів об’єктів, але й визначення їх структури, правильного виділення частин об’єктів у зображенні. Повністю згорткові нейронні мережі, такі як AlexNet-FCN [1] та SegNet [2] можуть здійснювати ефективну сегментацію зображень.

В роботі пропонується використати середовище навчання глибоких нейронних мереж з використанням GPU – DIGITS (The NVIDIA Deep Learning GPU Training System). Візуальний веб-інтерфейс DIGITS та потужні можливості дозволяють дослідникам спростити задачу навчання глибоких нейронних мереж. DIGITS може використовуватися для швидкого навчання глибоких нейронних мереж для задач класифікації зображень, сегментації зображень та виявлення об'єктів у зображеннях [3].

Метою роботи є дослідження семантичної сегментації зображень при використанні згорткових нейронних мереж та середовища DIGITS.

Результати дослідження

Для першого дослідження для навчання нейронної мережі було використано базу зображень PASCAL VOC [4]. Особливостями бази є наявність розміченого набору зображень для задачі сегментації та наявність 20 розмічених класів об’єктів у зображеннях.

В процесі навчання власних моделей у середовищі DIGITS найкращим набором параметрів навчання виявився такий: Solver type — Stochastic gradient descent, наперед навчена модель — AlexNet-FCN, policy (функція зміни кроку навчання) — Step Down (Step size=33, Gamma=0.1), base

learning rate (базовий крок навчання) — 0.0001, кількість епох навчання — 50. На рисунку 2 зображено результат тестування моделі на базі PASCAL VOC.

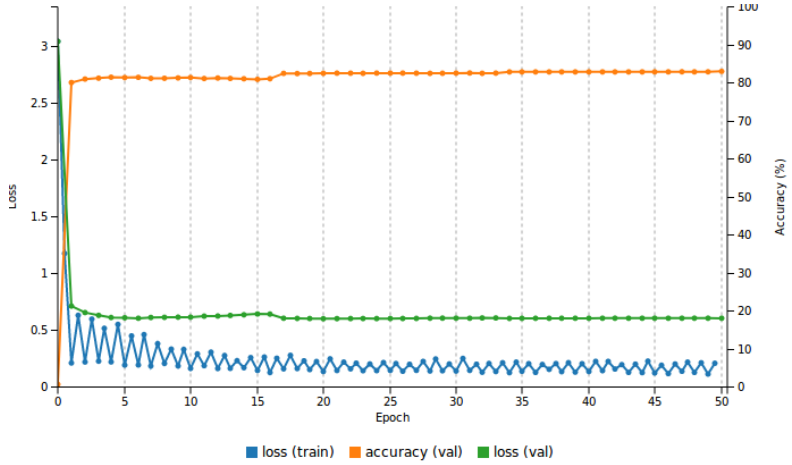

Рис. 1. Результати тестування моделі на базі PASCAL VOC

Як видно з рисунку 1 в результаті навчання моделі на тестових зображеннях отримано вірогідність сегментації близько 83%, що досить не погано для 20 класів об’єктів. На рисунку 2 зображено приклад роботи моделі на базі PASCAL VOC.

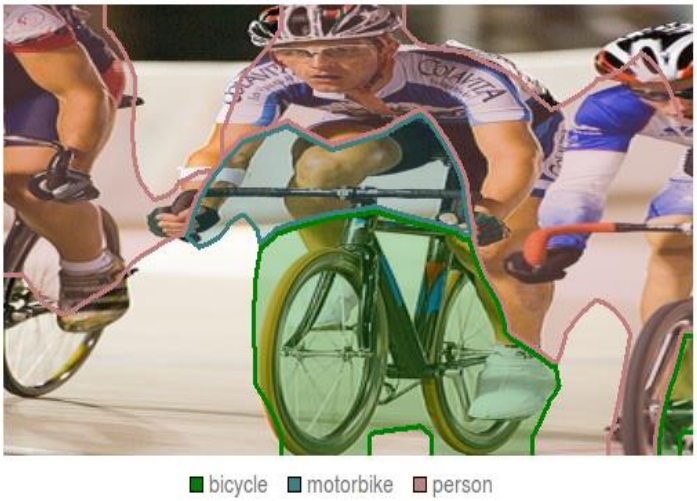

Рис. 2. Приклад роботи моделі на базі PASCAL VOC

Для другого дослідження для навчання нейронної мережі було використано базу зображень NVIDIA-Aerial Drone Dataset [4]. Зображення з цієї бази представляють собою аерофотозображення отримані з дронів на висоті декількох сотень метрів.

В процесі навчання моделі визначено оптимальні параметри навчання: Solver type — Stochastic gradient descent, наперед навчена модель — AlexNet-FCN, policy (функція зміни кроку навчання) — Exponention decay (Gamma=0.97), base learning rate (базовий крок навчання) — 0.0001, кількість епох навчання — 30. На рисунку 3 зображено результат роботи меделі на базі NVIDIA-Aerial Drone Dataset.

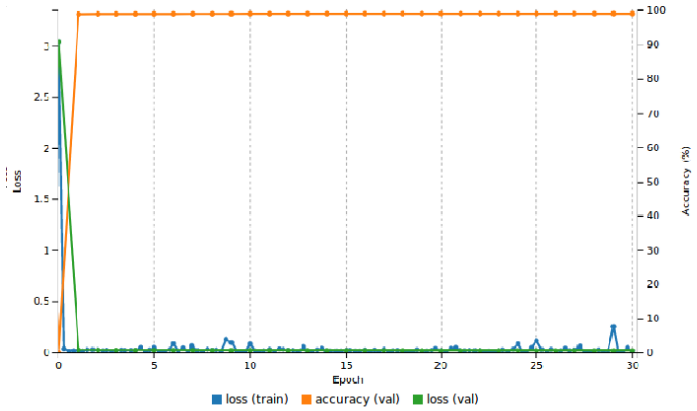

Рис. 3. Результати тестування моделі на базі NVIDIA-Aerial Drone Dataset

Як видно з рисунку 3 під час навчання власної моделі при використанні тестових зображень бази NVIDIA-Aerial Drone Dataset отримано дуже високу вірогідність сегментації близько 99%. Це пояснюється меншою кількістю класів в порівнянні з базою VOC, лише два класи - “небо” та “земля”. На аерозображеннях, отриманих з дронів у місті Вінниці якість сегментації при використанні власної моделі була дещо меншою, але все ж достатньою високою (рис. 4).

Рис. 4. Приклад застосування моделі навченій на базі NVIDIA-Aerial Drone Dataset для сегментації фото міста Вінниця

Висновки

В роботі проведені дослідження задачі семантичної сегментації зображень при використання середовища DIGITS. Для створення власних моделей у середовищі DIGITS використана нейромережева модель AlexNet-FCN та бази зображень PASCAL VOC та NVIDIA-Aerial Drone Dataset. В результаті навчання було досягнуто вірогідності сегментації у 83% (для тестового набору бази PASCAL VOC) та 99% (для тестового набору бази NVIDIA-Aerial Drone Dataset).

Розглянутий підхід є перспективним для подальших досліджень семантичної сегментації.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Long J. Fully Convolutional Networks for Semantic Segmentation [Електронний ресурс]: berkeley.edu. – Режим доступу: https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf.

2. Badrinarayanan V. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling [Електронний ресурс] / V. Badrinarayanan, A. Handa, and R. Cipolla // arXiv: 1511.00561, 2015. - Режим доступу: https://arxiv.org/pdf/1511.00561.pdf.

3. NVIDIA DIGITS [Електронний ресурс] – Режим доступу до ресурсу: https://developer.nvidia.com/digits. – Назва з екрану.
4. The PASCAL Visual Object Classes [Електронний ресурс]: robots.ox.ac.uk. Режим доступу: http://host.robots.ox.ac.uk/pascal/VOC/.
5. NVIDIA-aerial drone dataset [Електронний ресурс]: nvidia.box.com - Режим доступу: https://nvidia.box.com/shared/static/ft9cc5yjvrbhkh07wcivu5ji9zola6i1.gz.

Бездітний Михайло Васильович — студент групи 1АКІТ-17м, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, м. Вінниця, e-mail: Mykhailo_Beaditnyi@Jabil.com
Маслій Роман Васильович— к.т.н. доцент кафедри АІІТ, Вінницький національний технічний університет, м. Вінниця.

Bezditnyi Mykhailo V. — Faculty of Computer Systems and Automatics,, Vinnytsia National Technical University, Vinnytsia, email : Mykhailo_Bezditnyi@Jabil.com
Maslyi Roman V. — Ph.D., Associate Professor, Department of Automation and Intelligent Information Technologies, innytsia National Technical University, Vinnytsia