

МАТВІЙЧУК ЯРОСЛАВ

Національний університет «Львівська політехніка»

ORCID: [0000-0002-5570-182X](https://orcid.org/0000-0002-5570-182X)[yaroslav.m.matviychuk@lpnu.ua](mailto:yaroslav.m.matviychuk@lpnu.ua)

ЯЦИШИН ВОЛОДИМИР

Національний університет «Львівська політехніка»

ORCID: [0009-0001-9727-5080](https://orcid.org/0009-0001-9727-5080)[volodymyr.p.yatsyshyn@lpnu.ua](mailto:volodymyr.p.yatsyshyn@lpnu.ua)

## РОЗРОБЛЕННЯ АРХІТЕКТУРИ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ ВІЙСЬКОВОЇ ТЕХНІКИ НА ОБМЕЖЕНОМУ НАБОРІ ДАНИХ

Робота спрямована на розробку архітектури згорткової нейронної мережі для класифікації зображень військової техніки. Ключовою вимогою до моделі є можливість навчання на обмежених вибірках даних. Було взято існуючу модель VGG-16, в якій замінено класифікатор на власний, що базується на повністю зв'язній нейронній мережі з 2-а виходами. Усі натреновані моделі перевіряються на адекватність та в подальшому будуть використовуватися для швидкої ідентифікації військової техніки у потоці відео.

**Ключові слова:** згорткова нейронна мережа, класифікація зображень, модель машинного навчання, ідентифікація військової техніки

MATVIYCHUK YAROSLAV, YACISHYN VOLODYMYR  
Lviv Polytechnic National University

## DEVELOPMENT OF A CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE FOR MILITARY EQUIPMENT IMAGE CLASSIFICATION ON A LIMITED DATASET

The work is aimed at developing the architecture of a convolutional neural network for the classification of military equipment images. The key requirement for the model is the ability to be trained on limited data samples. The existing VGG-16 model was chosen, in which fully connected layers were replaced with a classifier based on a fully connected neural network with 2 outputs, resulting in a model with 13 convolutional blocks with a maximizing aggregation layer between each and 3 fully connected layers. Each convolutional layer was pre-trained using the ReLU activation function. Two fully connected layers of the replaced classifier also use ReLU as an activator, the last one-node layer uses a sigmoid function to perform the classification. In order to prevent retraining of the network, the Dropout regularization method with a screening factor of 0.2 was applied. To train the convolutional neural network, we use the Normal vs Military Vehicles dataset. The size of this dataset is quite limited and consists of approximately 17,500 files divided into training, verification and testing sets, each being divided in two asset classes: military and non-military. To overcome this limitation, the transfer learning method was used, when elements of a previously trained model are reused in a new machine learning model. Since VGG-16 is trained on a very large image set and is specially designed for image recognition and classification, resulting models show high accuracy and performance even after being trained on a limited datasets. In our case showing a decent 82% accuracy on Normal vs Military Vehicles validation dataset. All trained models are checked for adequacy, and will be used in the future for quick identification and classification of military equipment in the video stream.

**Keywords:** convolutional neural network, image classification, machine learning model, identification of military equipment

### Вступ

Завдання класифікації зображень полягає у тому, щоб на основі вхідного зображення визначити клас або ймовірність належності до класу, який найкраще описує зображення. Для людей таке вміння розпізнавання є однією із перших базових навичок. Коли людина бачить зображення або спостерігає навколишній світ, у більшості випадків може відразу визначити сцену і присвоїти кожному об'єкту категорію. Ці навички швидкого розпізнавання шаблонів, узагальнення попередніх знань і адаптації до різних зображень ще не повністю доступні машинам. Подібним чином комп'ютер може виконувати класифікацію зображень, шукаючи ознаки низького рівня, такі як краї та криві, а потім створювати більш абстрактні поняття за допомогою серії згорткових шарів. Це загальна ідея роботи згорткової нейронної мережі.

### Аналіз літературних джерел

Згорткові нейронні мережі (ЗНМ) є частиною сімейства глибоких нейронних мереж прямого поширення, де проміжний шар отримує на вхід ознаки, вилучені попереднім рівнем та передає свої вихідні дані наступному шару. Згорткові нейронні мережі працюють на основі фільтрів, які розпізнають певні характеристики зображення, наприклад, прямі лінії, прості кольори, вигини тощо [1].

Ядро фільтра – це матриця чисел (ваг). Ядро "навчається" з метою пошуку на зображеннях певних характеристик. Група ядер утворює фільтр. Кожен фільтр є ідентифікатором ознаки.

Фільтр рухається вздовж зображення і визначає, чи є певна потрібна характеристика в конкретній його частині. Для отримання відповіді працює операція згортки, яка є сумою добутків елементів фільтра та матриці вхідних сигналів. Площа, яку охоплює фільтр (рис. 2), називається сприйнятливим (рецептивним) полем. Сприйнятливим полем – частина матриці вхідних сигналів, що піддається згортці. Варто зазначити, що в міру заглиблення в мережу сприйнятливим полем фільтра збільшується, розширюється.

Фільтр переміщується вздовж матриці вхідних сигналів з певним кроком (stride). Крок визначає: на скільки пікселів зміщується фільтр. Якщо крок рівний, наприклад, одиниці, фільтр ковзатиме картою ознак піксель за пікселем. Це є типовим випадком.

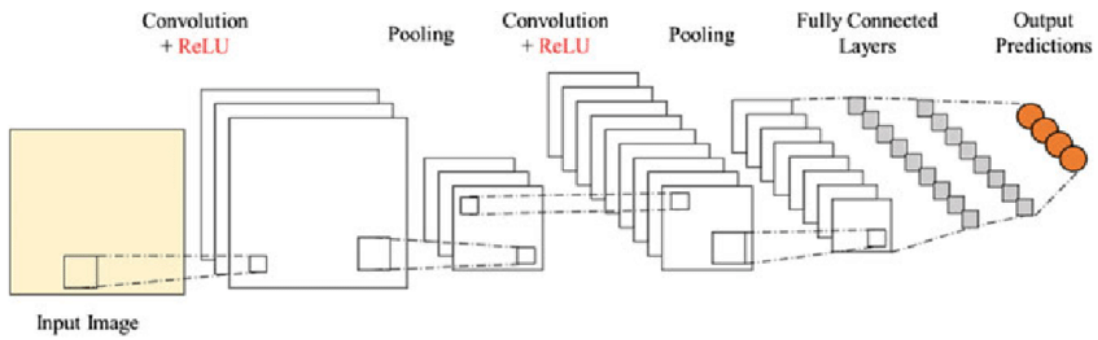


Рис. 1. Типова архітектура ЗНМ

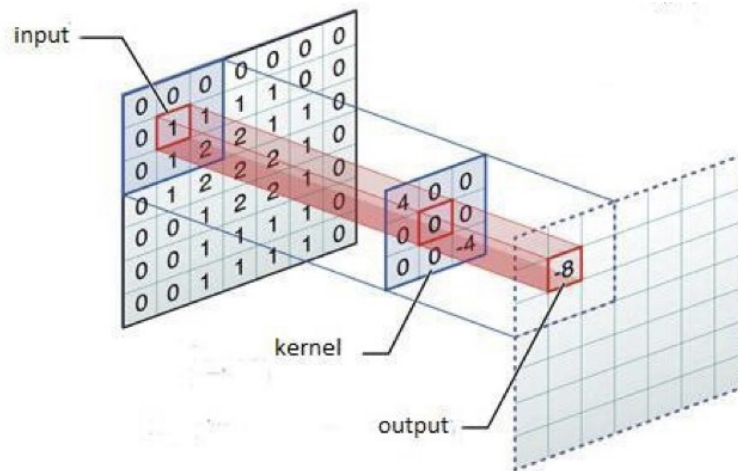


Рис. 2. Результат застосування фільтру

Якщо певна характеристика присутня у фрагменті зображення, охопленому фільтром, операція згортки на виході видаватиме число з відносно великим значенням. Якщо ж характеристика відсутня, вихідне число буде невеликим. На рис. 3 наведено конкретний приклад згортки [2, 4].

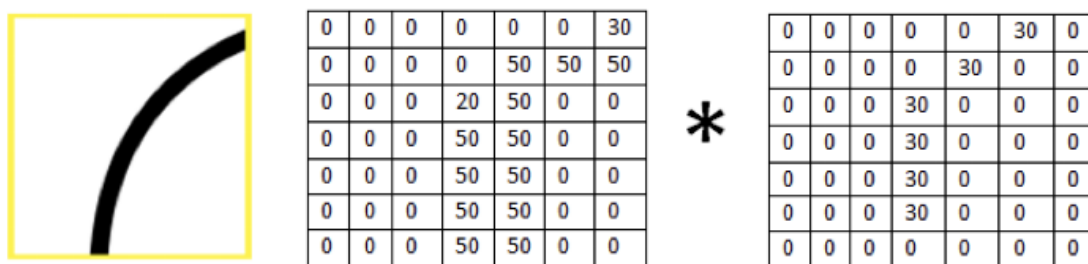


Рис. 3. Фільтр «шукає» лівосторонні криві

Якщо вхідними даними для першого згорткового шару є оригінальне зображення, то на вхід другого шару подається вихід першого шару у вигляді активаційних карт ознак, сформованих на попередньому етапі згортки. Таким чином, вихід шару визначає місця на вхідному зображенні, де є певні елементи низького рівня. Далі на виході наступного набору фільтрів будуть активаційні карти, які відповідають функціям вищого рівня (півкола, квадрати тощо). Послідовність згорткових шарів формує активаційні карти ознак, які представляють все більш складні функції.

З аналізу існуючих рішень визначено мету роботи – побудувати архітектуру системи ідентифікації військових транспортних засобів, що може бути натренована на обмеженому наборі даних.

**Основний матеріал**

Для навчання нейронної мережі скористаємось набором даних Normal vs Military Vehicles [5]. Розмір цього набору даних є досить обмеженим і складається із приблизно 17500 файлів у двох класах зображень: військова та не військова техніка. Структура наведена на рис. 4.

Для подолання обмеження скористаємось методом Transfer learning – перенесення навчання, коли

елементи попередньо натренованої моделі повторно використовуються в новій моделі машинного навчання. За основу візьмемо блоки згортки моделі VGG-16 [1], що була представлена на щорічному конкурсі ImageNet Large Scale Visual Recognition Challenge (ILSVRC). VGG16 – це згорткова нейронна мережа, навчена на підмножині набору даних ImageNet, колекції з понад 14 мільйонів зображень, що належать до 22 000 категорій. Ця модель запропонована у [2]. Якщо відкинути повністю зв'язні шари, то мережа складається з 13 згорткових шарів, має понад 14 мільйонів натренованих параметрів і завершується шаром максимального агрегування. Як новий класифікатор, у [2] додано два повністю пов'язані шари з кількістю нейронів 128 кожен, а також вихідний шар з одним вузлом (бінарна класифікація). З метою запобігання перенавчання мережі було застосовано метод регуляризації Dropout [3] з коефіцієнтом відсіювання 0.2. Повна характеристика одержаної мережі наведена на рис. 5.

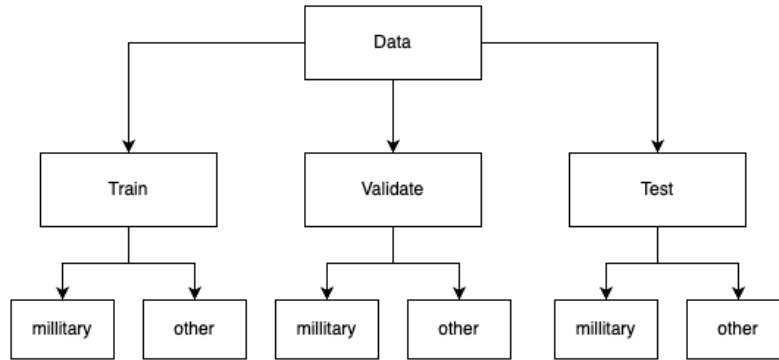


Рис. 4. Структура набору даних Normal vs Military Vehicles

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[None, 224, 224, 3]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 128)	3211392
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 128)	16512
dense_2 (Dense)	(None, 1)	129
Total params: 17,942,721		
Trainable params: 3,228,033		
Non-trainable params: 14,714,688		

Рис. 5. Параметри отриманої моделі

На рис. 6. наведено класифікаційний звіт після оцінки моделі на валідаційній вибірці. Враховуючи обмежений розмір набору даних, модель досягла доволі хорошого результату – 82% на валідаційних даних.

	precision	recall	f1-score	support
military	0.8	0.86	0.83	198
other	0.85	0.79	0.82	198
accuracy			0.82	396
macro avg	0.82	0.82	0.82	396
weighted avg	0.82	0.82	0.82	396

Рис. 6. Класифікаційний звіт після оцінки моделі на валідаційній вибірці

### Висновки

Для розв'язання задачі ідентифікації військового транспортного засобу було використано попередньо натреновану глибоку нейронну мережу VGG-16 за принципом методу перенесення навчання та здійснено заміну повністю пов'язаних шарів на власний класифікатор. Для побудови моделі було використано відкриту нейромережеву бібліотеку Keras, що є надбудовою до відкритого фреймворку для глибокого навчання TensorFlow.

Виходячи з того, що доступний набір даних налічує обмежену кількість зображень, для даної задачі було прийнято рішення використовувати попередньо натреновану модель VGG-16. Було здійснено заміну повністю зв'язаних шарів, в результаті чого модель налічувала 13 згорткових блоків з максимізаційним агрегаційним шаром між кожним з них та 3 повністю пов'язані шари. Кожен згортковий шар був попередньо натренований з використанням функції активації ReLU. Два повністю пов'язані шари заміненого класифікатора також використовують ReLU в ролі активатора, останній шар з одним вузлом використовує сигмоїдну функцію для здійснення класифікації.

Наступними завданнями дослідження можна визначити тренування і верифікацію моделі на більш спеціалізованих наборах даних.

### References

1. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images (2019). International Journal of Scientific and Research Publications, p. 9420. DOI: 10.29322/IJSRP.9.10.2019.p9420
2. Simonyan Karen, Zisserman Andrew (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition.
3. Dropout in (Deep) Machine learning. <https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5>
4. Javed F.M., Shamrat Mehedi, Chakraborty Sovon (2021). A Model Based on Convolutional Neural Network (CNN) for Vehicle Classification. Conference Paper CIS. [https://www.researchgate.net/publication/363634204\\_A\\_Model\\_Based\\_on\\_Convolutional\\_Neural\\_Network\\_CNN\\_for\\_Vehicle\\_Classification](https://www.researchgate.net/publication/363634204_A_Model_Based_on_Convolutional_Neural_Network_CNN_for_Vehicle_Classification)
5. Normal vs Military Vehicles. <https://www.kaggle.com/datasets/amanrajbose/normal-vs-military-vehicles?resource=download>