
СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

УДК 004.93

М. С. МАМУТА, І. В. КРАВЧЕНКО, О. Д. МАМУТА, С. Є. ТУЖАНСЬКИЙ

ОЦІНКА КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ ДЛЯ ПЕРЕНЕСЕННЯ НАВЧАННЯ У ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖАХ

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» 03056, пр-т. Перемоги 37, м. Київ, Україна, E-mail: rybalkomaryna@gmail.com

Інститут фізики Національної академії наук України, пр-т Науки 46, м. Київ, Україна, E-mail: mamuta.aleksandr@gmail.com

Вінницький національний технічний університет 21021, Хмельницьке шосе 95, м. Вінниця, Україна, E-mail: slavat@vntu.edu.ua

Анотація. У статті представлено порівняльний аналіз основних згорткових нейронних мереж, що використовуються для задач класифікації зображень. Розглянуто особливості прийому перенесення навчання для задачі класифікації. Досліджено ефективність перенесення навчання при використанні основних типів мереж відкритої нейромережної бібліотеки Keras.

Ключові слова: нейронна мережа, комп'ютерний зір, глибинне навчання, перенесення навчання

Abstract. The article provides an overview of the main neural networks used for image classification tasks. Their main advantages and disadvantages are discussed. The peculiarities of transfer learning for classification are analyzed. The effectiveness of transfer learning when using the basic networks of the Keras library is examined.

Key words: neural network, computer vision, deep learning, transfer learning

DOI: 10.31649/1681-7893-2023-45-1-64-70

ВСТУП

Системи комп'ютерного зору з елементами штучного інтелекту сьогодні є невід'ємною частиною нашого життя. Стрімкий розвиток таких систем зумовлений удосконаленням технологій глибинних нейронних мереж, що здійснюють ефективну обробку зображень та відео.

Однією із основних задач комп'ютерного зору є задача класифікації зображень і відео. Для її вирішення застосовуються згорткові нейронні мережі (ЗНМ) [1]. Бурхливий розвиток ЗНМ розпочався з 2012 року, коли найкращою у змаганнях з розпізнавання зображень бази ImageNet (ILSVRC Challenge) була визнана мережа AlexNet [2]. А вже у 2015 році на ILSVRC Challenge 2015 мережа ResNet перевершила не тільки інші наявні методи, а й можливості людини, отримавши похибку у точності розпізнавання лише 3,6% (людина у середньому робить близько 5% помилок) [3].

Разом з тим основною складністю роботи глибинних ЗНМ, що обмежує їх широке застосування є тривале навчання, яке може складати від кількох днів до місяців. Крім того, для успішного навчання потрібна велика кількість розмічених даних. У зв'язку з цим крім побудови мережі з нуля (from scratch) активно застосовується підхід перенесення навчання (transfer learning) [4]. Для глибинних мереж використовується два основних підходи до перенесення навчання [4]:

- метод, що ґрунтується на моделі (model-based);
- метод, що ґрунтується на ознаках (feature-based).

Зазвичай обидва підходи застосовуються одночасно.

Актуальним є дослідження ефективності прийому перенесення навчання при використанні різних ЗНМ.

СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

1. ОСНОВНІ ЕЛЕМЕНТИ АРХІТЕКТУРИ ЗНМ

В загальному випадку архітектура ЗНМ складається з таких шарів (рис. 1) [5]:

- згорткові шари (convolution), які виділяють ознаки;
- згладжуючий шар, який перетворює тензори ознак у вектор (flatten layer);
- повнозв'язні шари (fully connected) або класифікатор.

Вхідне зображення подається на перший згортковий шар, в результаті чого отримують карту ознак (feature map). Ця карта містить прості, низькорівневі ознаки. Далі карта ознак подається на вхід другого згорткового шару, в результаті чого отримують нову карту ознак, де виділяються більш складні ознаки (ознаки, виділені на попередньому етапі, об'єднуються). Тобто зі збільшенням глибини (кількості) згорткових шарів виділяються більш складні, високорівневі ознаки. Завдяки цьому можливо ефективно розпізнавати приховані закономірності та виділяти комплексні образи на зображеннях.

Для виділення ознак на кожному згортковому шарі застосовується набір фільтрів – ядер (kernel) заданого розміру [1].

Після кожного згорткового шару застосовується функція активації [6]. В глибоких ЗНМ це зазвичай функція ReLU.

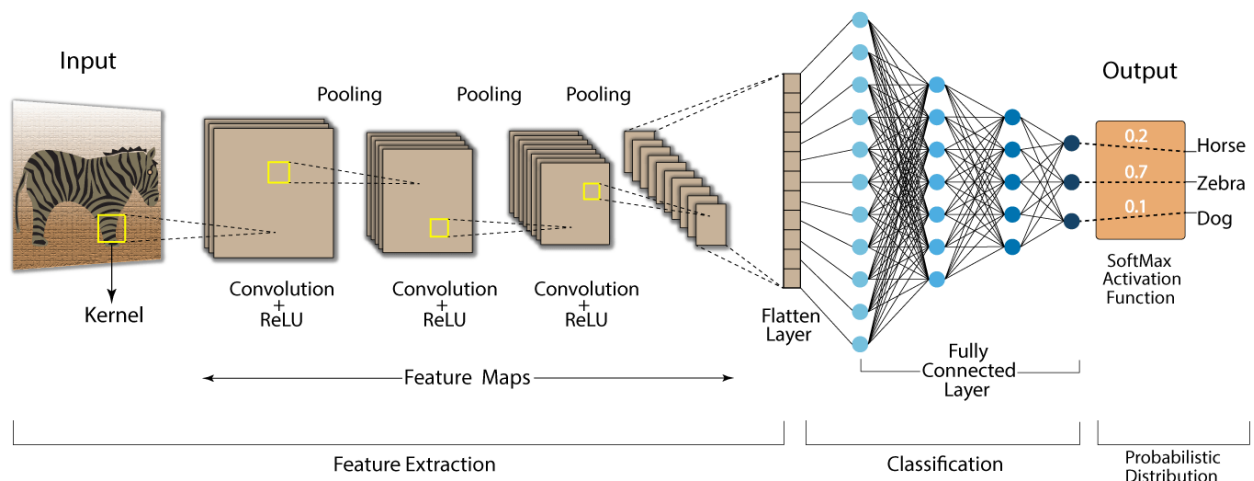


Рисунок 1 – Основні елементи архітектури ЗНМ для задачі класифікації зображень [5]

З метою зменшення розмірності карт та вибору найбільш значущих ознак застосовують операцію проріджування (pooling, частіше maxpooling) після кожного згорткового шару. Дана операція дозволяє зменшити кількість параметрів, а отже полегшити навчання.

Після останнього згорткового шару карти ознак витягуються у вектор та подаються на вхід класифікатора.

Кількість нейронів в останньому повнозв'язаному шарі ЗНМ має відповідати кількості класів, за якими здійснюється класифікація. Розрахунок імовірності належності об'єкта до того чи іншого класу проводиться функцією активації Softmax [6].

Для навчання ЗНМ зазвичай використовується алгоритм зворотного поширення помилки та метод градієнтного спуску.

Вхідними параметрами навчання ЗНМ – кількість та розмір фільтрів для кожного згорткового шару, функція активації, розмір фільтрів для проріджування, кількість нейронів та функція активації для повнозв'язних шарів, регуляризація вагових коефіцієнтів (в разі необхідності) [7].

Параметри, які визначаються в результаті навчання, – це вагові коефіцієнти повнозв'язних шарів та ядра фільтрів. Тобто мережа сама підбирає фільтри для виділення тих чи інших ознак на зображенні.

Основна проблема, яка характерна для ЗНМ та для всіх глибоких мереж, – це проблема затухання градієнта.

Причиною цього є те, що на початковому етапі вагові коефіцієнти задаються за допомогою генератора випадкових чисел, а отже вони ще є «поганими», і лише в процесі навчання ці коефіцієнти набувають сенсу. При використанні алгоритму зворотного поширення помилки потрібно градієнти передати з кінця на початок мережі, що викликає складність. Якщо градієнти мають невелике значення,

СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

то при великій кількості шарів вони перетворюються в нуль, тобто затухають. При цьому вагові коефіцієнти перших шарів майже не змінюються, тобто залишаються такими ж «поганими».

Ще однією складністю при роботі з нейронними мережами є необхідність великої кількості даних. Звичайно ж є готові набори, ще й розмічені, інколи навіть з результатами роботи стандартних алгоритмів. В такому разі зручно перевірити та порівняти роботу своєї моделі з існуючими. Але складності виникають при вузькоспеціалізованих задачах, оскільки набори даних зазвичай відсутні, принаймні у відкритому доступі.

2. АНАЛІЗ ЗНМ ДЛЯ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ

AlexNet - одна з перших ЗНМ з нескладною архітектурою.

AlexNet приймає на вхід кольорове зображення розміром 224×224 пікселі. Має 5 згорткових шарів та 3 повнозв'язних. Кількість нейронів в останньому повнозв'язному шарі – 1000, тобто класифікація здійснюється на 1000 класів.

Особливостями AlexNet, що дозволили їй перемогти в 2012 році на конкурсі ILSVRC, є:

- використання функції активації ReLU після кожного шару (крім останнього) замість гіперболічного тангенса та сигмоїди;
- паралельне навчання на двох GPU;
- аугментація даних.

Після успіху AlexNet були розроблені нові глибинні ЗНМ, основними з яких є VGG [8], Inception [9], ResNet та DenseNet [10].

VGG по своїй суті – також досить проста ЗНМ. Головна її відмінність від AlexNet – глибина, вона має більшу кількість згорткових та повнозв'язних шарів.

Ще однією особливістю є використання фільтрів розміром 3×3. У AlexNet на першому згортковому шарі застосовуються фільтри розміром 11×11. А отже VGG здатна виділяти більш локальні ознаки.

Кількість шарів у VGG залежить від виду мережі, існують VGG11, VGG13, VGG16, VGG19. Найчастіше використовуються VGG16 та VGG19 з кількістю шарів 16 та 19 відповідно, зокрема вони дають найкращий результат при роботі з базою ImageNet.

Особливістю ЗНМ GoogleNet є використання декількох фільтрів різних розмірів на кожному згортковому шарі (технологія inception). Таким чином, карта ознак одночасно сканується декількома фільтрами, які виділяють різну інформацію з вхідних даних. Далі здійснюється конкатенація отриманих результатів і передача в наступний шар.

Для боротьби в GoogleNet із затуханням градієнтів використовуються додаткові класифікатори, тобто блок повнозв'язних шарів розташовується не лише після всіх згорткових шарів, а й в декількох місцях всередині мережі. Функція втрат (loss) рахується для всіх класифікаторів і для алгоритму зворотнього поширення помилки використовується складний loss. За рахунок цього вся мережа ефективно навчається.

Особливістю ЗНМ ResNet є використання технології residual connection в боротьбі зі затуханням градієнта.

Існує багато варіантів ЗНМ ResNet, від ЗНМ ResNet-18, ЗНМ ResNet-34, ЗНМ ResNet-50 до ЗНМ ResNet-152. Звичайно ж ці мережі відрізняються не лише кількістю шарів, а й архітектурою блоків.

Цікавим фактом щодо мережі ResNet є те, що при видаленні декількох шарів мережа не втрачає якості. Цей ефект досягається завдяки технології residual connection, що дає можливість передавати інформацію через декілька шарів.

ЗНМ DenseNet складається з Dense блоків. Особливістю DenseNet є те, що вихід кожного шару подається не лише в наступний, а в усі наступні шари, тобто кожен наступний шар отримує всі карти ознак всіх попередніх шарів. Таким чином, кожен наступний шар отримує більше інформації про зображення. Всі карти ознак мають однаковий розмір, що досягається завдяки додатковим шарам, які називаються transition layer.

DenseNet має декілька переваг:

- strong gradient flow, тобто сигнал loss добре передається в усі шари, включаючи перші;
- мережа містить меншу кількість параметрів;
- в класифікатор поступають ознаки всіх рівнів складності;
- DenseNet добре працює з невеликими наборами вхідних даних (data set).

СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

3. ОЦІНКА КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ БІБЛІОТЕКИ KERAS

В комп'ютерному зорі для задачі класифікації зображень стандартним та ефективним підходом при глибокому навчанні на невеликій кількості даних є перенесення навчання (transfer learning). Він полягає в тому, що навчена нейронна мережа з відомою архітектурою використовується для вирішення нової задачі [6]. Процес відбувається в декілька кроків:

- імпортується навчена нейронна мережа з відомою архітектурою;
- виділяється згорткова частина, яка забезпечує виділення ознак, тобто не підключається класифікатор навченої мережі, а вагові коефіцієнти згорткової частини «заморожуються» і не приймають участі в навчанні;
- створюється і навчається новий класифікатор.

Таким чином, навчається лише та частина мережі, яка відповідає за класифікацію, що дозволяє значно пришвидшити процес навчання.

Якщо набір даних, на якому навчалась базова мережа значно відрізняється від нового набору даних поточної задачі, то доцільним є не лише навчання класифікатора, а й декількох шарів згорткової частини. Такий метод називається тонким налаштуванням (fine tuning), і є частиною transfer learning. Зазвичай навчаються більш глибокі згорткові шари, оскільки перші виділяють загальні, низькорівневі ознаки. Кількість шарів для заморожування залежить від відмінності між наборами даних, складності задачі та обсягу нового набору даних.

Для дослідження був використаний хмарний сервіс Google Colaboratory (Google Colab). Досліджено ефективність перенесення навчання для найбільш популярних нейронних мереж бібліотеки Keras – VGG16, InceptionV3 та ResNet50. Класифікація здійснювалась за шістьма класами зображень: літак, машина, військовий літак, військова машина, військовий гелікоптер, танк. Зображення взяті з мережі інтернет, з них сформований власний набір даних, який розбитий на три множини: тренувальну, валідаційну (для навчання) та тестову (для перевірки роботи). Кількість екземплярів в кожному класі – менше 100 одиниць.

Нижче наведені основні налаштування та результати ефективності (в якості критерію ефективності розглядаємо точність класифікації об'єктів) обраних мереж для бібліотек VGG16, InceptionV3 та ResNet50.

Розглянемо оцінку точності класифікації зображень для VGG16. Спершу здійснюється завантаження мережі:

```
from tensorflow.keras.applications import VGG16
```

Далі виділяємо згорткову частину. При цьому визначаються три аргументи:

- аргумент `weights` визначає джерело вагових коефіцієнтів для ініціалізації моделі;
- аргумент `include_top` визначає підключення класифікатора. За замовчуванням – це класифікатор на 1000 класів з бази ImageNet. При перенесенні навчання даний класифікатор не підключається;
- аргумент `input_shape` визначає форму тензорів, які будуть подаватись на вхід системи.

```
vgg16_net = VGG16(weights='imagenet', include_top=False,  
input_shape=input_shape)
```

Забороною навчання завантажених шарів мережі:

```
vgg16_net.trainable = False
```

Створюємо нову модель зі згортковою частиною з VGG16 та власним класифікатором:

```
model = Sequential()  
model.add(vgg16_net)  
model.add(Flatten())  
model.add(Dense(256, activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(len(classes), activation='softmax'))
```

Далі здійснюємо компіляцію моделі:

```
model.compile(loss='categorical_crossentropy',  
optimizer=Adam(learning_rate=1e-5),  
metrics=['accuracy'])
```

та її навчання з вказаною кількістю ітерацій (epoch):

```
history = model.fit(
```

СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

```
train_generator,  
steps_per_epoch=nb_train_samples // batch_size,  
epochs=10,  
validation_data=val_generator,  
validation_steps=nb_validation_samples // batch_size)
```

Перевірка на тестових даних показала точність 66,67%.

При застосуванні методу fine tuning було дозволено навчання лише останнього згорткового шару мережі VGG16, який називається "block5_conv3".

```
# Активуємо для навчання всю мережу  
vgg16_net.trainable = True  
# Перебираємо в циклі прошарки та вимикаємо навчання для всіх прошарків,  
окрім потрібного  
trainable = False  
for layer in vgg16_net.layers:  
    if layer.name == 'block5_conv3':  
        trainable = True  
layer.trainable = trainable
```

Перевірка на тестових зображеннях показала точність 73,33%.

Оцінимо точність класифікації для бібліотеки InceptionV3. Спершу здійснюємо завантаження мережі:

```
from tensorflow.keras.applications import InceptionV3
```

Подальші налаштування здійснюються аналогічно як і для мережі VGG16. При застосуванні власного класифікатора перевірка на тестових даних показує точність 86,67% вже після 5-ти епох навчання.

Для методу fine tuning навчання здійснюється починаючи з 249 шару згорткової частини:

```
for layer in model.layers[:249]:  
    layer.trainable = False  
for layer in model.layers[249:]:  
    layer.trainable = True
```

Перевірка на тестових даних показала точність 93,33%.

Оцінка точності класифікації ResNet50 здійснювалась таким чином. Спершу здійснюємо завантаження мережі:

```
from tensorflow.keras.applications.resnet50 import ResNet50
```

Подальші налаштування відбуваються аналогічно як і для мережі VGG16.

При застосуванні власного класифікатора перевірка на тестових даних показує точність 30%. Збільшення кількості епох не дає підвищення ефективності. Припускаємо, що мережа ResNet потребує більшої кількості вхідних даних.

Метод fine tuning покращення результатів також не демонструє.

ВИСНОВКИ

В статті здійснено порівняльний аналіз сучасних ЗНМ для задач класифікації зображень. Розглянуто особливості основних типів згорткових нейронних мереж відкритої бібліотеки Keras, які використовуються для задачі класифікації даних і зображень у системах комп'ютерного зору. Досліджено ефективність використання таких глибинних нейромереж при застосуванні техніки перенесення навчання. Встановлено, що найкращі результати при використанні перенесення навчання для невеликої кількості вхідних даних дає ЗНМ InceptionV3, точність при цьому сягає 93,33%.

СПИСОК ЛІТЕРАТУРИ

1. Yu. Tomka, M. Talakh, V. Dvorzhak, O. Ushenko, "Implementation of a Convolutional Neural Network Using TensorFlow Machine Learning Platform", Optoelectronic Information-Power Technologies, vol. 44, №2, pp. 55-65, 2023. DOI: <https://doi.org/10.31649/1681-7893-2022-44-2-55-65>

СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

2. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", [Electronic resource]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
3. K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", [Electronic resource]. Available: <https://arxiv.org/pdf/1512.03385.pdf>.
4. Qiang Yang, Yu Zhang, Wenyuan Dai, Sinno Jialin Pan. "Transfer Learning", Cambridge University Press, 2020, p. 380. DOI: <https://doi.org/10.1017/9781139061773>
5. Nafiz Shahriar, "What is Convolutional Neural Network – CNN (Deep Learning)", [Electronic resource]. Available: <https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn-deep-learning-b3921bdd82d5>
6. Ian Goodfellow, Yoshua Bengio, Aaron Courville. "Deep learning", MIT Press, 2017, p. 800
7. А. Момот, М. Мамута. "Інтелектуальні та інформаційні системи. Практикум", КПІ ім. Ігоря Сікорського, 2022, 221 с.
8. [Electronic resource]. Available: <https://ela.kpi.ua/bitstream/123456789/47696/1/IIS.pdf>
9. K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", [Electronic resource]. Available : <https://arxiv.org/pdf/1409.1556.pdf>
10. C. Szegedy. "Going Deeper with Convolutions", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9. DOI: 10.1109/CVPR.2015.7298594
11. K. He, X. Zhang, S. Ren, J. Sun. "Deep Residual Learning for Image Recognition", [Electronic resource]. Available: <https://arxiv.org/pdf/1512.03385.pdf>.

REFERENCES

1. Yu. Tomka, M. Talakh, V. Dvorzhak, O. Ushenko, "Implementation of a Convolutional Neural Network Using TensorFlow Machine Learning Platform", Optoelectronic Information-Power Technologies, vol. 44, №2, pp. 55-65, 2023. DOI: <https://doi.org/10.31649/1681-7893-2022-44-2-55-65>
2. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", [Electronic resource]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
3. K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", [Electronic resource]. Available: <https://arxiv.org/pdf/1512.03385.pdf>.
4. Qiang Yang, Yu Zhang, Wenyuan Dai, Sinno Jialin Pan. "Transfer Learning", Cambridge University Press, 2020, p. 380. DOI: <https://doi.org/10.1017/9781139061773>
5. Nafiz Shahriar, "What is Convolutional Neural Network – CNN (Deep Learning)", [Electronic resource]. Available: <https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn-deep-learning-b3921bdd82d5>
6. Ian Goodfellow, Yoshua Bengio, Aaron Courville. "Deep learning", MIT Press, 2017, p. 800
7. А. Момот, М. Мамута. "Інтелектуальні та інформаційні системи. Практикум", КПІ ім. Ігоря Сікорського, 2022, 221 с.
8. [Electronic resource]. Available: <https://ela.kpi.ua/bitstream/123456789/47696/1/IIS.pdf>
9. K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", [Electronic resource]. Available : <https://arxiv.org/pdf/1409.1556.pdf>
10. C. Szegedy. "Going Deeper with Convolutions", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9. DOI: 10.1109/CVPR.2015.7298594
11. K. He, X. Zhang, S. Ren, J. Sun. "Deep Residual Learning for Image Recognition", [Electronic resource]. Available: <https://arxiv.org/pdf/1512.03385.pdf>.

Надійшла до редакції 05.03.2023р.

СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

МАМУТА МАРИНА СЕРГІЇВНА – к.т.н., старший викладач кафедри комп'ютерно-інтегрованих оптичних та навігаційних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, *[e-mail: rybalkomaryna@gmail.com](mailto:rybalkomaryna@gmail.com)*

КРАВЧЕНКО ІГОР ВОЛОДИМИРОВИЧ – старший викладач кафедри комп'ютерно-інтегрованих оптичних та навігаційних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, *[e-mail: kravchenko.igor@iit.kpi.ua](mailto:kravchenko.igor@iit.kpi.ua)*

МАМУТА ОЛЕКСАНДР ДМИТРОВИЧ – к.т.н., науковий співробітник відділу когерентної та квантової оптики Інституту фізики Національної академії наук України, м. Київ, *[e-mail: mamuta.aleksandr@gmail.com](mailto:mamuta.aleksandr@gmail.com)*

ТУЖАНСЬКИЙ СТАНІСЛАВ ЄВГЕНОВИЧ – к.т.н., доцент кафедри біомедичної інженерії та оптико-електронних систем Вінницького національного технічного університету, м. Вінниця, *[e-mail: slavat@yntu.edu.ua](mailto:slavat@yntu.edu.ua)*

Maryna MAMUTA, Ihor KRAVCHENKO, Olexander MAMUTA, Stanyslav TUZHANSKY

EVALUATION OF IMAGE CLASSIFICATION FOR TRANSFER LEARNING IN CONVOLUTIONAL NEURAL NETWORKS

National Technical University of Ukraine "Ihor Sikorskyi Kyiv Polytechnic Institute"
03056, pr. Peremohy 37, Kyiv, Ukraine
Institute of Physics of the National Academy of Sciences of Ukraine, 46 Nauky Avenue, Kyiv, Ukraine
Vinnytsia National Technical University 21021, Khmelnytske Shosse 95, Vinnytsia, Ukraine