

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.9312

О.В. БУБЕНЩИКОВ, Є.В. ЛЕПА
Херсонський національний технічний університет**ВИКОРИСТАННЯ ЗГОРТАЛЬНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ІДЕНТИФІКАЦІЇ ОБЛИЧЧЯ ЛЮДИНИ**

В даній роботі розглянута нейронна мережа, завданням якої є ідентифікація обличчя людини. Ідентифікація особи людини зводиться до розв'язку завдання класифікації, а модель представлена у вигляді згортальної нейронної мережі. У якості прототипу взята нейронна мережа Lenet-5, яка застосовувана для ідентифікації рукописних цифр. Враховуючи, що ідентифікація особи є більш складним завданням, чим ідентифікація рукописних цифр, була змінена архітектура нейронної мережі.

Запропоновано три варіанти мережі різної архітектури, зроблене їхнє навчання на одній і тій же вибірці зображень обличчя людини, яка була створена для розв'язку завдання. При цьому використані різні функції активації й методи оцінки точності ідентифікації. Мінімізація функції помилки при ідентифікації виконана методами стохастичного градієнтного спуску (SGD) та Adam. У якості оцінки нейронних мереж різної архітектури використані час навчання й точність ідентифікації.

Перший варіант містить 4 шари згортки, 2 шари ЗНМ підвибірки (субдискретизації), вхідний шар та два повнозв'язних шари, один з яких є вихідним шаром. Час навчання та точність класифікації обличчя людини в цій мережі відповідно дорівнює 22,5 хвилин і 50%.

Другий варіант містить 3 шари згортки, 3 шари підвибірки, також вхідний шар та два повнозв'язних шари, один з яких є вихідним шаром. Час навчання та точність класифікації в цій мережі відповідно дорівнює 9,2 хвилин і 81,25%.

Найкращою, з точки зору точності класифікації та часу навчання, є третій варіант архітектури нейронної мережі. Вона побудована на основі другого варіанта мережі, але кількість нейронів в першому повнозв'язному шарі дорівнює 1024, замість 512. Час навчання та точність класифікації обличчя людини в цій мережі відповідно дорівнює 9,1 хвилин і 93,75%.

Подальше підвищення якості може бути досягнуте при збільшенні набору даних і, насамперед, вибірки навчання.

Ключові слова: класифікація, ідентифікація, нейронна мережа, згортальна нейронна мережа, моделювання, Lenet-5, шар субдискретизації, повнозв'язний шар, вихідний шар.

А.В. БУБЕНЩИКОВ, Е.В. ЛЕПА
Херсонский национальный технический университет**ИСПОЛЬЗОВАНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ИДЕНТИФИКАЦИИ ЛИЦА ЧЕЛОВЕКА**

В данной работе рассмотрена нейронная сеть, задачей которой является идентификация лица человека. Идентификация личности человека сводится к решению задачи классификации, а модель представлена в виде сверточной нейронной сети. В качестве прототипа взята нейронная сеть Lenet-5, которая применяется для идентификации рукописных цифр. Учитывая, что идентификация личности является более сложной задачей, чем идентификация рукописных цифр, была изменена архитектура нейронной сети.

Предложено три варианта сети различной архитектуры, реализовано их обучение на одной и той же выборке изображений лиц людей, которая была создана для решения данной задачи. При этом использованы различные функции активации и методы оценки точности идентификации. Минимизация функции ошибки при идентификации выполнена методами стохастического градиентного спуска (SGD) и Adam. В качестве оценки нейронных сетей различной архитектуры использованы время обучения и точность идентификации.

Первый вариант содержит 4 слоя свертки, 2 слоя СНС подвыборки (субдискретизации), входной слой и два полносвязных слоя, один из которых является выходным слоем. Время обучения и точность классификации лица человека в этой сети соответственно равна 22,5 минут и 50%.

Второй вариант содержит 3 слоя свертки, 3 слоя подвыборки, также входной слой и два полносвязных слоя, один из которых является выходным слоем. Время обучения и точность классификации в этой сети соответственно равна 9,2 минут и 81,25%.

Лучшей, с точки зрения точности классификации и времени обучения, является третий вариант архитектуры нейронной сети. Она построена на основе второго варианта сети, но количество нейронов в первом полносвязном слое равно 1024, вместо 512. Время обучения и точность классификации лица человека в этой сети соответственно равна 9,1 минут и 93,75%.

Дальнейшее повышение качества может быть достигнуто при увеличении набора данных и, прежде всего, выборки обучения.

Ключевые слова: классификация, идентификация, нейронная сеть, сверточная нейронная сеть, моделирование, Lenet-5, слой субдискретизации, полносвязный слой, выходной слой.

O.V. BUBENSHCHYKOV, E.V. LIERA
Kherson National Technical University

USE OF CONVOLUTIONAL NEURAL NETWORKS FOR HUMAN FACE IDENTIFICATION

In this work was considered a neural network, whose task is to identify a person's face. Identification of a person's personality is reduced to the solution of the classification problem, and the model is represented as a convolutional neural network. Lenet-5 neural network is used, as a prototype, to identify handwritten digits. Considering that facial identification is a more complex task than identifying handwritten digits, the architecture of the neural network has been changed.

Three variants of the network of different architecture was proposed, they trained on the same sample of images of a person, which was created for solving the problem. In this case, lots of activation functions and methods for assessing the accuracy of the identification was used. Minimization of the error function for identification was performed using stochastic gradient descent (SGD) and Adam methods. As an estimation of neural networks of different architecture, time of training and accuracy of identification was used.

The first variant contains 4 layers of convolution, 2 layers of CNN sub-sampling, an input layer and two full layers, one of which is the output layer. Training time and classification accuracy of a person's face in this network are 22.5 minutes and 50%.

The second variant contains 3 layers of convolution, 3 layers of the sub-sample, an input layer and two full layers, one of which is the output layer. The time of training and the accuracy of classification in this network are 9.2 minutes and 81.25%.

The third variant is the best option for the architecture of the neural network because of the accuracy of the classification and the training time. It is based on the second version of the network, but the number of neurons in the first full layer is 1024, instead of 512. The training time and the accuracy of the classification of a person's face in this network are 9.1 minutes and 93.75%.

Further quality improvement can be achieved by increasing the set of data and, the training sample.

Key words: classification, identification, neural network, convolutional neural network, modeling, Lenet-5, sub-sampling layer, full layer, output layer.

Постановка проблеми

Останнім часом системи ідентифікації людини набирають все більшої популярності наприклад, для доступу до персональних комп'ютерів, смартфонів, в сфері безпеки різних установ. Однією з основних є система ідентифікації по зображенню людини. До переваг даної системи можна віднести простоту і мобільність обладнання, а також масовість ідентифікації. Ідентифікація зображення зводиться до розв'язку задачі багатомірної класифікації.

Для вирішення задачі розпізнавання осіб існує безліч методів і алгоритмів, серед яких можна виділити підходи, засновані на нейронних мережах, на гнучкому порівнянні графів, на прихованих Марковських моделях, на розкладанні Кархунена-Лоева, лініях однакової інтенсивності, і т.д [1].

Аналіз останніх досліджень і публікацій

Аналіз методів роботи з зображеннями показав, що для вирішення даного завдання ефективніше буде використовувати штучні нейронні мережі (ШНМ), в зв'язку з тим, що вони забезпечують можливість отримання класифікатора, що добре моделює складну функцію розподілу зображень обличчя. Основною перевагою використання нейронних мереж для виявлення та ідентифікації особи, є здатність до навчання даної системи для виділення ключових характеристик особи з навчальних наборів даних.

Формулювання мети дослідження

Метою роботи була побудова моделі знань у вигляді нейронної мережі для ідентифікації обличчя людини, з оптимальною структурою, яка забезпечувала б максимальну якість ідентифікації. Як основа, була використана відома архітектура згортальної нейронної мережі Lenet-5 [3].

Викладення основного матеріалу дослідження

ЗНМ навчається зроблених наборах даних, що містять по п'ятдесят зображень десяти людей, яких треба класифікувати. Ці зображення заздалегідь оброблені, та представлені у форматі JPG, у розмірі 225x150. Крім того, є п'ятдесят зображень інших людей для формування класу людини, що не міститься в базі. Зображення зчитується входами нейронної мережі, а один з виходів потрібен для виведення результату (ім'я людини).

В якості основи для ЗНМ для класифікації обличчя людини було обрано архітектуру LeNet-5, оскільки вона є досить простою в реалізації, і видає високий відсоток точності при вирішенні задач класифікації зображень. Приклад такої архітектури для класифікації рукописних цифр, представлено на рис. 1.

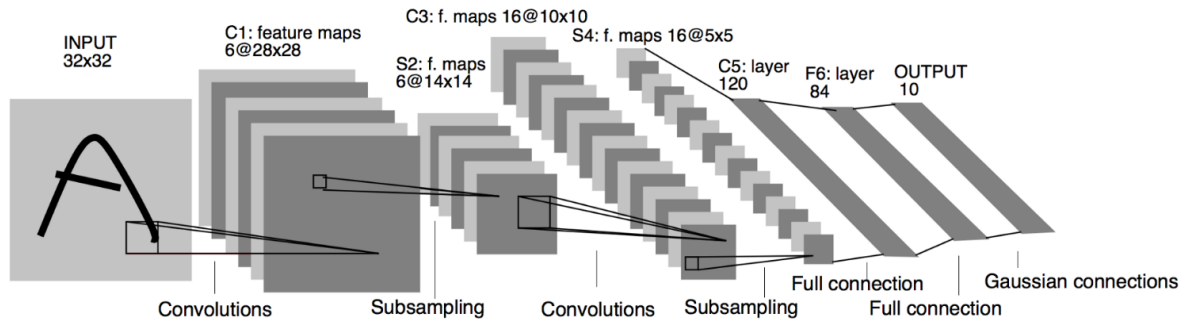


Рис. 1. Архітектура ЗНМ для класифікації рукописних цифр

Ідея цієї архітектури полягає в чергуванні шарів згортки (Convolutions) і шарів субдискретизації (Subsampling) для виявлення ключових характеристик на зображеннях та наявності повнозв'язних шарів (Full connection) на виході для класифікації зображень.

Якщо розглядати процес більш детально, то можна зазначити, що шар згортки включає в себе для кожного каналу свій фільтр та ядро згортки, яке обробляє попередній шар за фрагментами, підсумовуючи результати матричної похідної для кожного фрагмента. Вагові коефіцієнти ядра згортки невідомі і встановлюються в процесі навчання. Шар субдискретизації являє собою нелінійне ущільнення карти ознак, при цьому група пікселів (зазвичай розміру 2x2) ущільнюється до одного пікселя, проходячи нелінійне перетворення. Після кількох проходжень згортки і ущільнення за допомогою субдискретизації, дані об'єднуються і передаються вже на звичайну повнозв'язну нейронну мережу.

Але використання цієї ЗНМ в початковому вигляді не є доцільним, оскільки архітектура цієї ЗНМ направлена більше на класифікацію цифр. Зображення обличчя людини є більше складним і потребує більш складної архітектури нейронної мережі, за рахунок зміни кількості шарів згортки та субдискретизації для підвищення знаходження кількості ключових характеристик на зображенні.

Тому було змінено структуру ЗНМ, шляхом додання шарів згортки та субдискретизації. Крім того, було змінено вхідний шар, що пов'язано з розміром зображення, який в результаті експериментів було визначено рівнем 225x150 у трьох градаціях RGB (червоний-зелений-блакитний). Якщо вибрати розмір занадто маленький, то мережа не зможе виявити ключові ознаки осіб, якщо розмір буде занадто великий, то обчислювальна складність мережі підвищиться. Збільшено також кількість нейронів на повнозв'язному шарі, оскільки збільшується кількість ключових характеристик. Останній повнозв'язний шар є шаром-класифікатором, та видає відсоток того, яка людина зображена на вхідному зображенні. В результаті було отримано перший варіант архітектури мережі, яка представлена на рис. 2.

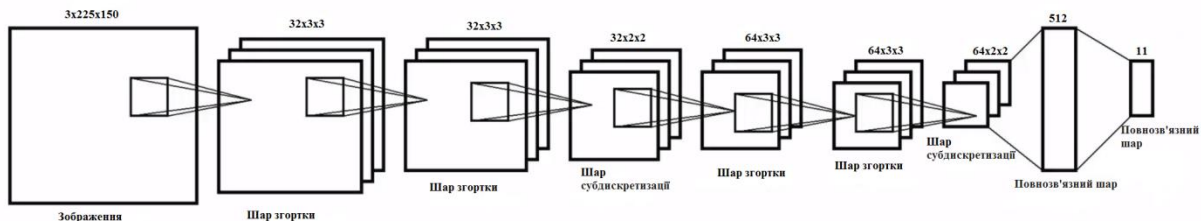


Рис. 2. Архітектура першого варіанту мережі

Згідно цієї архітектури ЗНМ складається з 9 шарів. Перший шар є вхідним та містить 101250 нейрони ($225 \times 150 \times 3$). Другий та третій шари є згортальними та містять по 32 карти ознак в кожному шарі. Карта ознак – це набір шарів нейронів, кожен з яких використовує різні ядра згортки, для пошуку різних признаков на зображенні. Ядро являє собою фільтр або вікно, яке ковзає по всій області попередньої карти і знаходить певні ознаки об'єктів розміром 3×3 . Розмір ядра зазвичай визначається від ситуації. Якщо розмір ядра дуже маленький, то воно не зможе виділити будь-які ознаки, якщо занадто великий, то збільшується кількість зв'язків між нейронами.

Четвертий шар є шаром субдискретизації складається з 32 карт ознак і ядром 2×2 . Шостий та п'ятий шари – згортальні з 64 картами ознак у кожному з ядром 3×3 . Сьомий шар – шар субдискретизації також з 64 картами ознак та ядром 2×2 . Восьмий шар – повнозв'язний шар, розміром в 512 нейронів. Останній дев'ятий шар – повнозв'язний вихідний шар з 11 нейронами, кожен з яких позначає ту чи іншу людину з бази даних, або її відсутність у ній.

Таким чином, на першому та другому шарах згортки існує 864 ($32 * (3 \times 3) * 3$) зв'язки у кожному, на першому шарі дискретизації – 384 ($32 * (2 \times 2) * 3$). На третьому та четвертому шарах згортки – 1728 ($64 * (3 \times 3) * 3$) зв'язків, на другому шарі дискретизації – 768 ($64 * (2 \times 2) * 3$). На першому та другому повнозв'язних шарах – 5632 ($512 * 11$) зв'язків. У результаті нейронна мережа має 11968 зв'язків.

Наступним етапом після побудови ЗНМ іде навчання цієї мережі. Для цього використовується вибірка навчання з образів обличчя людей. На вхід нейронної мережі подаються образи й ваги зв'язків між нейронами настроюються так, щоб на виході мережі отримані образи як найменше відрізнялися б від тих, які вже класифіковані у вибірці навчання.

Від результатів навчання нейронної мережі буде залежати якість розв'язку завдання ідентифікації обличчя людини. Оцінка результатів навчання може бути зроблена за наступними показниками:

- швидкість навчання;
- кількість епох навчання;
- функції активації;
- функції мінімізації помилки.

Для того щоб можна було почати навчання нейронної мережі потрібно визначитися з тим, як вимірювати якість розпізнавання зображення. Для цього можна використовувати найпоширенішу в теорії нейронних мереж функцію середньоквадратичної помилки [5].

$$E^p = \frac{1}{2} (D^p - O(I^p W))^2, \quad (1)$$

де E^p - це помилка розпізнавання для p -ї навчальної пари,

D^p - бажаний вихід мережі,

$O(I^p, W)$ - вихід мережі, що залежить від p -го входу і вагових коефіцієнтів W , куди входять ядра згортки, зміщення, вагові коефіцієнти S - і F - шарів.

Завдання навчання полягає у визначенні такого налаштування ваг W , щоб вони для будь-якої навчальної пари (I^p, D^p) давали мінімальну помилку E^p . Щоб порахувати помилку для всієї навчальної вибірки береться середнє арифметичне по помилках для всіх навчальних пар.

В процесі експериментів з ЗНМ було виявлено, що для мінімізації функції помилки E^p , в контексті рішення задачі класифікації зображень, найефективнішими є метод стохастичного градієнтного спуску (SGD) [7] та метод Adam [8].

Метод стохастичного градієнтного спуску оновлює кожен параметр, віднімаючи градієнт функції, яка оптимізується по відповідному параметру. Потім ться його на швидкість навчання. Однак метод SGD може розходитися, якщо швидкість навчання налаштована невірно. Метод адаптивної інерції (Adam) змінює масштаб швидкості навчання для кожного параметра окремо, з урахуванням всіх минулих градієнтів для цього. Це ефективно зменшує крок навчання для параметрів, які мають велику величину градієнта [8].

Одним з найважливіших аспектів нейронних мереж є функція активації (activation function), яка привносить в мережі нелінійність, яка робить їх універсальними апроксиматорами функцій. Функція активації - це спосіб нормалізації вхідних даних. Тобто, якщо на вхід ШНМ подається велике число, то пропустивши його через функцію активації, можна отримати вихід в потрібному діапазоні. Найбільш часто використовуються наступні функції активації [10]:

1. Лінійна функція - найпростіша з функцій активації. Має вихід такий же, як і вхід.
2. Порогова функція активації - проста лінійна функція, яка часто використовується в нейронних мережах. Результат приймає значення 0 для негативного аргументу і 1 для позитивного аргументу.

3. Сигмоїдальна функція - монотонно зростаюча нелінійна функція з насиченням. Сигмоїд дозволяє підсилювати слабкі сигнали і не насичуватись від сильних сигналів. Прикладами сигмоїдальної функції активації може служити логістична функція або гіперболічний тангенс.

4. Rectified linear unit (ReLU) або «випрямляч» (rectifier, за аналогією з однополуперіодним випрямлячем в електротехніці) є найбільш часто використовуваною функцією активації. Функція визначається наступною формулою [9]

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (2)$$

Функція активації ReLU активно використовуються в глибоких нейронних мережах для задач комп'ютерного зору і розпізнаванні мови. Її застосування істотно підвищує швидкість збіжності стохастичного градієнтного спуску, в деяких випадках до 6 разів. Також функція активації ReLU не схильна до насичення і вимагає виконання менш ресурсномістких операцій в порівнянні з сигмоїдою та гіперболічним тангенсом. Тому було в якості функції активації в шарах згортки обрано функцію RELU.

На вихідному шарі використовується функцію активації Softmax [11]. Цю функцію доцільно використовувати в задачах класифікації, оскільки вона представляє собою зважену і нормовану на одиницю суму експонент. Якщо вхідні дані представляють собою вибірку з будь-якого експоненціального розподілу, то виходи елементів можна трактувати як ймовірності, що підходить для рішення поставленої задачі.

Навчання нейронної мережі було проведено на процесорі Intel Core i7-4500 CPU @ 1.80 GHz. Результати навчання мережі першого варіанту архітектури ЗНМ (рис. 2), представлені у табл. 1.

Таблиця 1

Характеристики навчання

Функція мінімізації помилки	Кількість епох навчання	Швидкість навчання (lr)	Час навчання (хвилин)	Точність класифікації, %
SGD	25	0.1	14,6	25
Adam	25	0.1	16,6	35
SGD	30	0.01	20	37
Adam	30	0.01	22,5	50
SGD	35	0.02	23,4	34
Adam	35	0.02	26,3	45

В процесі експериментів було знайдено оптимальну величину для швидкості навчання ЗНМ = 0,01 lr. Кількість епох навчання знаходиться в діапазоні від 25 до 35. При значному збільшенні кількості епох починається процес перенавчання ЗНМ, при якому мережа добре класифікує дані з навчальної вибірки, але погано класифікує нові дані, які не брали участі в навчанні. При недостатній кількості епох, також зменшується точність класифікації.

Цій варіант архітектури ЗНМ забезпечує точність класифікації на рівні 45-50%. Для підвищення якості класифікації була змінена архітектура мережі (другий варіант), яка наведена на рис. 3.

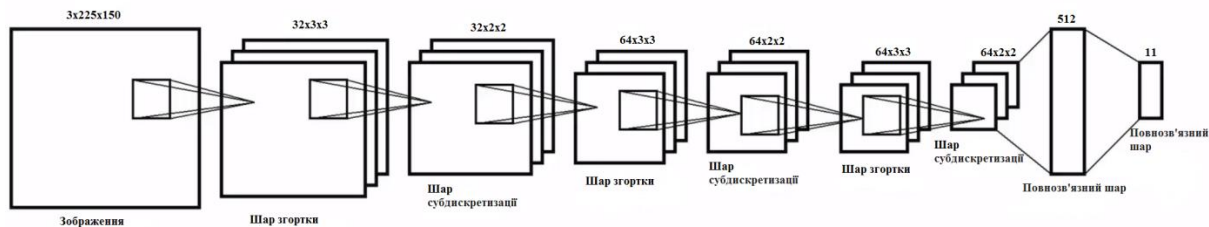


Рис. 3. Архітектура другого варіанту мережі

В нейронній мережі було вилучено третій та п'ятий шари згортки. Додано шар згортки з 64 картами ознак та ядром 3x3, шар субдискретизації також з 64 картами ознак, але з ядром 2x2 перед першим повноз'язним шаром.

Характеристики навчання нейронної мережі нової архітектури представлено у табл. 2.

Таблиця 2

Характеристики навчання				
Функція мінімізації помилки	Кількість епох навчання	Швидкість навчання (lr)	Час навчання (хвилин)	Точність класифікації, %
SGD	30	0.01	8,4	75
Adam	30	0.01	9,2	81,25
SGD	25	0.01	6,7	62,50
Adam	25	0.01	7,4	56,25
SGD	35	0.01	11,6	75
Adam	35	0.01	12,8	68,75

Точність класифікації збільшилась практично вдвоє до 81%. Також приблизно у два рази зменшився час навчання мережі.

Для подальшого підвищення точності класифікації була змінена архітектура нейронної мережі в такий спосіб. Збільшена кількість нейронів в першому повнозв'язному шарі з 512 на 1024. Це в свою чергу збільшить кількість міжнейронних зв'язків, що може позитивно відобразитись на класифікаційних здібностях цієї мережі. Наступну архітектуру мережі (третій варіант), представлено на рис. 4.

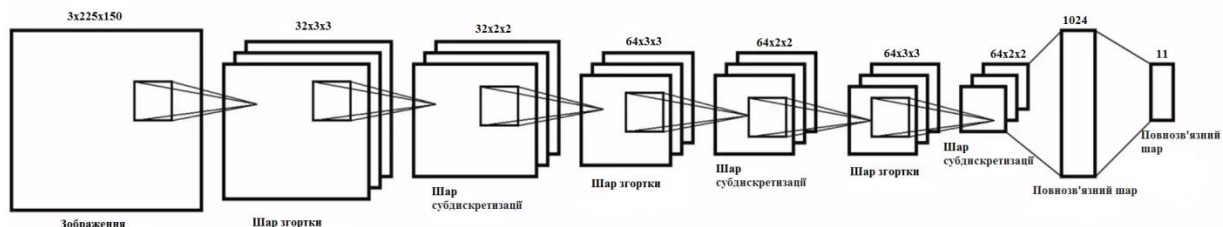


Рис. 4. Архітектура третього варіанту мережі

Характеристики навчання нейронної мережі зміненої архітектури представлено у табл. 3.

Таблиця 3

Характеристики навчання				
Функція мінімізації помилки	Кількість епох навчання	Швидкість навчання (lr)	Час навчання (хвилин)	Точність класифікації, %
SGD	30	0.01	8,5	87,6
Adam	30	0.01	9,1	93,75
SGD	25	0.01	6,7	84
Adam	25	0.01	7,5	82
SGD	35	0.01	11,9	84
Adam	35	0.01	12,7	91,3

Точність класифікації виросла до 94%, а всі інші характеристики практично не змінилися, у тому числі й час навчання. Таким чином, використовуючи досить просту архітектуру нейронної мережі, отримана висока точність класифікації.

Висновки

1. Виконано створення і навчання згортальних нейронних мереж, здатних реалізувати завдання класифікації обличчя людини.

2. Наведені основні етапи створення та навчання розглянутих нейронних мереж. В результаті було отримано три варіанта архітектури нейронної мережі, кожний з яких складається з 9 шарів.

Перший варіант містить 4 шари згортки, 2 шари ЗНМ підвибірки (субдискретизації), вхідний шар та два повнозв'язних шари, один з яких є вихідним шаром. Час навчання та точність класифікації обличчя людини в цій мережі відповідно дорівнює 22,5 хвилин і 50%.

Другий варіант містить 3 шари згортки, 3 шари підвибірки, також вхідний шар та два повнозв'язних шари, один з яких є вихідним шаром. Час навчання та точність класифікації в цій мережі відповідно дорівнює 9,2 хвилин і 81,25%.

3. Найкращою, з точки зору точності класифікації та часу навчання, є третій варіант архітектури нейронної мережі. Вона побудована на основі другого варіанта мережі, але кількість нейронів в першому

повнозв'язному шарі дорівнює 1024, замість 512. Час навчання та точність класифікації обличчя людини в цій мережі відповідно дорівнює 9,1 хвилин і 93,75%.

4. Подальше підвищення якості може бути досягнуте при збільшенні набору даних і, насамперед, вибірки навчання.

Список використаної літератури

1. Habr.com [електронний ресурс] Аналіз підходів розпізнання обличчя - <https://habr.com/company/synesis/blog/238129/>.
2. Habr.com [електронний ресурс] Опис особливостей згортальних нейронних мереж - <https://habr.com/post/309508/>.
3. Deep Learning [електронний ресурс] Опис архітектури Lenet-5 - <http://deeplearning.net/tutorial/lenet.html>.
4. ИНТУИТ [електронний ресурс] Визначення понять навчання НМ з вчителем та без вчителя - <https://www.intuit.ru/studies/courses/88/88/lecture/20555>.
5. Efficient BackProp [електронний ресурс] Опис функції середньоквадратической помилки - <http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf>.
6. Habr.com [електронний ресурс] Опис функції перехрестної ентропії - <https://habr.com/company/wunderfund/blog/314872>.
7. Machinelearning.ru [електронний ресурс] Опис методу градієнтного спуску - http://www.machinelearning.ru/wiki/index.php?title=Метод_градиентного_спуска
8. Habr.com [електронний ресурс] Опис функції мінімізації помилки Adam - <https://habr.com/post/318970>.
9. DataReview.info [електронний ресурс] Опис функції активації RELU - <http://datareview.info/article/eto-nuzhno-znat-klyuchevyie-rekomendatsii-po-glubokomu-obucheniyu-chast-2>.
10. AIPortal.ru [електронний ресурс] Функції активації в нейронних мережах - <http://www.aiportal.ru/articles/neural-networks/activation-function.html>.
11. Ievbras.ru [електронний ресурс] Опис функції активації Softmax - <http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content394/Content394.htm>