

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РК**

**АО "КАЗАХСТАНСКО-БРИТАНСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ"**

**МАГИСТЕРСКАЯ ПРОГРАММА**

**"ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ"**

**ОПИСАНИЕ СИСТЕМЫ**

**НА ТЕМУ:**

**DEPLOYING DEEP LEARNING ALGORITHMS**

**FOR IMAGE CLASSIFICATION**

**ВЫПОЛНИЛ: БЕКГАНБЕТОВ А.Р., ID: 16MD0136**

**РУКОВОДИТЕЛЬ: АКЖАЛОВА А.Ж., PH.D.**

**АЛМАТЫ**

**ДЕКАБРЬ, 2017**

**CONTENTS**

[**1. GENERAL INFORMATION**](#_gjdgxs) **2**

[1.1. Formulation of the problem](#_spa2op3judaf) 2

[1.2. Programming language and libraries](#_fr27q7y8vwrj) 2

[1.3. Normative references](#_tbv3o12zuvel) 2

[1.4. Terms, definitions and abbreviations](#_ji78rqs7zhd3) 2

[**2. FUNCTIONAL PURPOSE**](#_h98vewgcj2f3) **3**

[**3. DESCRIPTION OF THE LOGICAL STRUCTURE**](#_n06johv6rqwd) **4**

[3.1. Algorithm of the program](#_pc0ad9t2oax0) 4

[3.2. Implementation of the algorithm](#_9ahs6x303lj) 4

[3.3. Model specifications](#_23x2qyedgu1c) 5

[**CONCLUSION**](#_2tvmegureijp) **9**

[**APPLICATION**](#_2jxsxqh) **10**

# 1. GENERAL INFORMATION

This document is a description of the algorithm for the classification of images to be developed, describes the algorithm of actions, the logical structure and deployment of the application, determines the technical means used, as well as input and output data. The Document provides information on the task posed, the proposed algorithm and direct implementation, as well as other information that makes the technical and analytical aspects of the algorithm understandable and transparent.

## 1.1. Formulation of the problem

The task of classification is a formalized problem in which there are a number of objects (situations) separated in some way into classes. A finite set of objects is defined for which it is known which classes they belong to. This set is called sampling. The class affiliation of the remaining objects is unknown. It is required to construct an algorithm capable of classifying (see below) an arbitrary object from the original set.

To classify an object means to indicate the number (or name) of the class to which the object belongs.

Classification of the object - the number or class name, issued by the classification algorithm as a result of its application to this particular object.

In mathematical statistics classification problems are also called discriminant analysis problems. In machine learning, the classification problem is solved, in particular, using the methods of artificial neural networks in setting up an experiment in the form of training with the teacher.

## 1.2. Programming language and libraries

To implement the algorithm, the language "Python" is used.

To implement the manipulation of data libraries "numpy", "pandas", "keras", "tensorflow" are used.

For the drawing of graphs, the library "matplotlib", "seaborn" is used.

For ready-made solutions such as CNN or RandomForest, "keras" was used.

## 1.3. Normative references

The document is developed in accordance with the standards:

|  |  |
| --- | --- |
| Standard 19.402-2000 | «USPD. Program description» |

## 1.4. Terms, definitions and abbreviations

In machine learning, a *convolutional neural network* (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery.

CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics.

ConvNets derive their name from the “*convolution*” operator. The primary purpose of Convolution in case of a ConvNet is to extract features from the input image. Convolution preserves the spatial relationship between pixels by learning image features using small squares of input data.

# 2. FUNCTIONAL PURPOSE

The algorithm being developed is a tool for classification images . The tasks of categorization (classification) potentially solve the following tasks:

* Image recognition
* Face recognition
* Video analysis
* Natural language processing
* Drug discovery
* Checkers game learning
* Automatic translation from images

This algorithm is probably not an integrated solution, but under certain conditions and ease of implementation it can act as a temporary solution. Although, in the future we will see that in certain problems, this algorithm provides a good result.

The most successful applications of algorithm are:

* Crime image recognition
* Text translation
* Data tagging

It should be noted that this algorithm also offers a solution for the hierarchical structure of categories.

# 3. DESCRIPTION OF THE LOGICAL STRUCTURE

The algorithm for input accepts the classified (training) data, on the basis of which we build the model of the solution. Each sample data, we alternately add to the model. It is worth noting that the order of addition does not matter in our case. For each image encountered in the database, we calculate its share in each of the categories. In the future, turning to direct categorization, we use these values ​​to calculate the image contribution to each of the categories. Thus, if the image occurs uniformly in each of the categories, then the contribution will be the same. Below we consider the steps of the algorithm in more detail.

## 3.1. Algorithm of the program

1) Data extraction.

We have collected image data from the olx.kz. The total number of collected image is 560 000. We decided to split the whole dataset into train, validate and test parts in proportion of 60% / 20% / 20%. Thus, we have 336 000 images in train dataset, and 100 000 images in each of validate and test.

2) Data preparation.

1. The first step in data processing is to bring input images to the one common size. We should always remember that image size directly affects the memory and training time of the model.
2. Data augmentation means increasing the number of data points. In terms of images, it may mean that increasing the number of images in the dataset. In terms of traditional row/column format data, it means increasing the number of rows or objects.

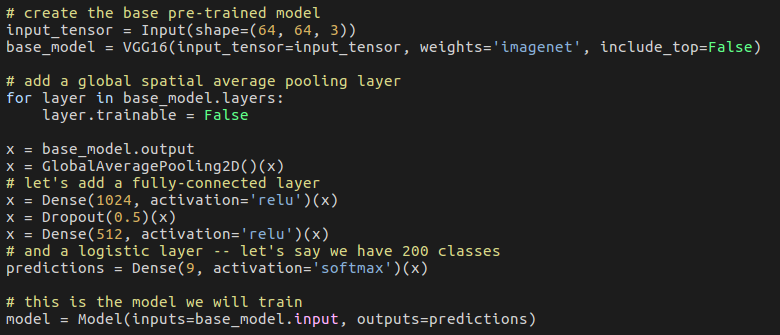
3) Model training

1. Model training is the step of optimization, in out case it’s optimization of the neural network with specific image that we have prepared for training. In our case different convolution neural network architectures are used to obtain results.
2. Testing this model on validation data. Getting results and picking model that satisfies requirements.

## 3.2. Implementation of the algorithm

One of the advantages of this algorithm is that it is easily implemented in any programming language. This paper presents an implementation in Python using only libraries for convenient data representation (numpy and pandas), which in turn is fine with almost all the now known tools and Python libraries.

Below is an example of code that implements model architecture and transfer learning technics.



*Figure 1. Transfer learning architecture*

In image 3.2.1 transfer learning technique was used. Transfer learning or inductive transfer is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.

## 3.3. Model specifications

The characteristics of the algorithm can be represented in the following aspects:



*Figure 2. Confusion matrix of the model on test data*

A confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). In figure 2 there is shown that some classes like animals are very easy to identify and some of the classes like services are crosses classes like animals, Fashion & Style.

|  |  |  |
| --- | --- | --- |
| **Model type** | **Model structure** | **Accuracy on test** |
| LeNet like | 1 convolutional layer | 19,5 |
|  | 3 convolutional layer | 29 |
|  | 3 convolutional layer, 1 fully connected layer | 33,5 |
|  | 3 convolutional layer, 2 fully connected layer | 37,3 |
| AlexNet like | 3 convolutional layer, 3 fully connected layer | 43,4 |
| VGG16 Transfer learning | Full transer learning, 12 convolutional, 2 fully | 44,1 |
|  | Partly transer learning, 12 convolutional, 3 fully, dropout layer | 45,1 |
| InceptionResNetV2  Transer learning | Full transer learning, 1 fully connected layer | 63,9 |
| Human accuracy (mean) |  | 68,1 |
| Human accuracy (top) |  | 75 |

*Figure 3. Table of results of the different architectures*

Figure 3 shows models that is trained.”Model type” column shows basic information of the architecture of the neural network, “Model” column contains information about it characteristics and “Test Accuracy” column show result on the Test dataset. The last 2 rows is the results of the humans. Data for this is collected via chatbot in 2 weeks and stored in database.

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Precision** | **Recall** | **F1-score** |
| Услуги | 44 | 40 | 42 |
| Хобби, отдых | 38 | 50 | 43 |
| Электроника | 80 | 33 | 47 |
| Недвижимость | 64 | 60 | 62 |
| Мода и стиль | 47 | 78 | 59 |
| Детский мир | 78 | 58 | 67 |
| Транспорт | 67 | 83 | 74 |
| Животные | 83 | 83 | 83 |
| Дом и сад | 86 | 86 | 86 |

*Figure 4 Precision, recall and f1-score on the test dataset. Result of the best model - “InceptionResNetV2”*

Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances, while recall (also known as sensitivity) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. Both precision and recall are therefore based on an understanding and measure of relevance. F1 score is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score.

# CONCLUSION

In this work we are reviewed several convolutional neural network architectures, implemented and tested it, deployed to the server, created clients to the model and compared human prediction and tested it.

Firstly, we started by collecting data and creating web crawler for it on Scrapy Framework. Data is stored in form of file database and images storage.

Secondly, all images are reshaped, normalized and optimized for storing in the HDF5 data format.

Thirdly, most common convolutional neural networks are reviewed. It has been found that convolutional neural networks is a highly challenging subject to understand and implement. In this step it was needed to find right combination of parameters: activation functions, size of kernels, padding sizes, tuning of hyper-parameters and neural network architecture.

Fourthly, we implemented backend server for the model serving, RESTful API and telegram client for it. In this step we collected data from users, analyzed our previous work. Model’s accuracy is nearly 63.9% is close to the human 68.1%.

In the future we plan to introduce the following items:

* Relational and nonrelational databases for storage and multi worker crawler;
* Recurrent neural networks to text data;
* More advanced convolutional neural network architectures;
* Extend classes, by adding subclasses;
* Improve dataset by adding advanced augmentation;

# APPLICATION

Below are all the necessary links and materials presented in the paper:

* <https://www.olx.kz/> - data OLX
* https://github.com/abaybek/master\_thesis - storage repository of all implementations and algorithm