

# Creating catalogues of clothes images using neural networks

Anna V. Korobko<sup>1</sup>, Aleksei A. Korobko<sup>2</sup> and Aleksei V. Markovin<sup>3</sup>

<sup>1</sup>*Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russian Federation*

<sup>2</sup>*Institute of Computational Modeling, SB RAS, Krasnoyarsk, Russian Federation*

<sup>3</sup>*"Osnova" Ltd., Krasnoyarsk, Russian Federation*

## Abstract

A lot of businessmen, companies and brands have created their accounts on Instagram using this social network as a platform for the promotion and sales of their goods, work and services. Despite all the possibilities which Instagram holds, the promotion of goods under the conditions of market saturation is a complicated task. It is becoming urgent to search for new technological solutions which would provide collecting images of goods from Instagram business accounts and aggregating this information in one integrated on-line marketplace, taking into account the requirements formed by the sales of clothes in e-commerce. The present study is devoted to the practical test of the approach to automatically cataloguing of goods based on images with the help of neural networks in the frames of collecting and aggregating information about goods from Instagram business accounts in an integrated on-line marketplace. The experience of applying neural network models is studied for the fashion industry in general and, in particular, for the cataloguing of clothes images. The applied methods and approaches of building convolutional neural networks are described and substantiated. The architecture of two neural network models for determining the colour and category of clothes from their images is described in detail. The accuracy of the model as well as the losses during learning and testing is estimated. The accuracy of the models is compared with the accuracy of a random classification. Testing the basic configurations allows one to determine the directions for future research, to formulate forthcoming scientific and technical problems and to form reference values of the classification accuracy for estimating the efficiency of more complex models.

## Keywords

processing, cataloguing, neural networks, fashion, Instagram

## 1. Introduction

With the development of information technologies, Internet is becoming an important aspect in the life of many people. This is unlimited access to knowledge, possibility of earning money remotely, communication with friends all around the world as well as a marketplace. Instagram is a social network based on a relatively new way of communication, i.e. uploading images and short videos. Instagram, which started from one million users in 2010 has at present more than 500 million active users who daily watch the information uploaded in the network.

---


*SibDATA 2021: The 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications 2021, June 25, 2021, Krasnoyarsk, Russia*

✉ [gglhroom@gmail.com](mailto:gglhroom@gmail.com) (A. V. Korobko)

🆔 0000-0001-5337-3247 (A. V. Korobko)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Many businessmen, companies and brands have created their accounts on Instagram using this network as a platform for the promotion and sales of their goods, work and services. This is an attractive marketplace due to numerous active users, unique possibilities for advertising, integration with Facebook and convenient tools of business analytics. However, in spite of all the possibilities of the network Instagram, the promotion of goods under the conditions of the saturated market is a difficult task. Most accounts which represent small and medium business remain unnoticed and have difficulties in attracting new users. A search for new technological solutions is becoming urgent in order to provide the collection of images of goods from business accounts of Instagram and aggregation of this information in an integrated marketplace taking into account the requirements formed by the sales of clothes in e-commerce.

The market of e-commerce has been developing at a high rate since 2013. At the end of 2019 in Russia there were about 4 700 clothes and footwear shops on the Internet with the level of sales being no lower than one delivery order per day; more than 100 shops are among the top 1000 leaders of the Russian e-commerce market. Customers already have a great experience of buying in e-commerce, and the requirements to presenting goods on the site of an internet shop have been formed. One can hardly imagine a marketplace or an internet clothes shop without a catalogue of goods. The most necessary attributes of goods which form the structure of the whole product line are “Colour” and “Category”.

The present study is devoted to the practical test of the approach to automatically cataloguing of goods based on images with the help of neural networks in the frames of collecting and aggregating information about goods from Instagram business accounts in an integrated on-line marketplace. The experience of applying neural net models is studied for the fashion industry in general and, in particular, for cataloguing clothes images. The applied methods and approaches of building convolutional neural networks are described and substantiated. The architecture of two neural network models for determining the colour and category of clothes from their images is described in detail. The accuracy of the model as well as the losses during learning and testing is estimated. The accuracy of the models is compared with the accuracy of a random classification. Testing the basic configurations allows one to indicate the directions for future research, to formulate forthcoming scientific and technical problems and to form reference values of the classification accuracy for estimating the efficiency of more complex models.

## **2. Review of the existing solutions**

At present, the most effective technique of classifying images is artificial neural networks. A neural net is a simple mathematical model which is not programmed but learns itself [1, 2, 3]. The model analyses a great number of examples related to the problem being solved, and finds in them statistical regularities to be used to form rules for automatically solving the stated problem. This approach is radically different from the earlier algorithms of artificial intellect, which required the preliminary formulation of the human knowledge. The use of a neural net both automatizes the process of problem-solving and the process of acquiring the knowledge which is necessary for it.

The accumulated experience of using neural networks for solving various problems of decision-making support allows forming a great number of parameters for tuning the model and

their possible values. These parameters include: architecture of the neural network, number and type of the layers, number of free neurons in each layer, set of input and target values, activation functions, algorithms and optimization functions, learning style and number of iterations. With this range of possible “tunings” of the neural network, the problem of constructing a model for solving a specific problem becomes a time-consuming procedure.

In English publications concerning the solution of the problems of fashion industry, use is made of the technologies of clothes image analysis, including neural networks. In the study [4] the problem of predicting a colour of clothes which will be fashionable in the next season is solved. In [Wang et al., 2018] a hybrid intellectual model of the medium term forecasts of the volume of sales in retail fashion trade is built. An approach is proposed to solve the problem of determining the clothes style which will become a new fashionable trend based on the image analysis in social networks [5].

The problem of cataloguing clothes which is considered in this study is most often formulated in literature as the classification of images based on the clothes type. In the study [6] one of the tasks of computer vision is solved, i.e. to determine clothes items in images. Under consideration are such items as: hat, glasses, bag, trousers, shoes, etc. The suggested approach is based on applying a sequence of modern methods: highlighting possible areas containing the object and using a convolutional neural network of deep learning. Since the location of items is in strong correlation with the location of human joints, the authors take into account the information about the stance of the man in the images in order to increase the efficiency of the algorithm. A qualitative and quantitative analysis of several models of convolutional networks based on 70 000 clothes images from the dataset Fashion-MNIST was made in [7]. An approach to using the semantic segmentation of clothes images as the preliminary stage of recognizing the clothes category is described in [8]. A model architecture is suggested including a basic network which serves for acquiring the features and a net in the form of the feature pyramid which sets the grouping of the feature values.

The review of the scientific literature and existing solutions shows that neural nets are intensively used for predicting fashion trends and automatic cataloguing of clothes images. The proposed solutions differ, depending on the research goal and list of the clothes items and accessories being considered. In the frames of the present study, it is interesting to test the basic neural net models with convolutional layers using our own dataset obtained from business accounts of the social network Instagram.

### **3. Methods and approaches**

The general approach to solving the stated problem is based on the methodology of the system analysis, object-oriented approach and theory of database. For testing the neural network models it is decided to use the language Python, a modern high-level programming language of general purpose oriented towards increasing the productivity of the developer and improving the readability of the code. Due to the human-readable program code and to a great number of in-built functional libraries, Python is highly suitable for computational experiments, scientific research and fast prototyping.

As a program environment for the implementation of the algorithms of identifying the

colour of items from a fly page, as well as their distribution and cataloguing, use was made of Jupyter, a web platform for the interactive development of executive pages (notebook), which can simultaneously contain information modules structured using the markup language Markdown, connected fragments of the program code (not only in Python) and results of the execution of this code. The flexibility of this development environment in combination with the high capacities in Python allows one to efficiently solve research problems. In the present research, use is made of a library for the support of processing big multidimensional arrays and matrices with a set of high-level mathematical functions (numpy), library of quick functions for analysing and manipulating the data based on the relational structure of representing multi-index datasets (pandas), library of high-level operations with files and collections of files (shutil), complete library for the creation of static, animated and interactive visualization (matplotlib) and framework for the deep machine learning of neural networks (keras).

Using the algorithm of collecting information and an assembly module for collecting images from sources (profiles of users) allowed accumulating a database containing 1253 images of goods from more than 150 accounts. The uploaded images were divided by the moderator into 23 categories and into 19 groups according to the colour of the proposed item. The problem of cataloguing of goods is reduced to a mathematical problem of single label multi-class classification. This means that each of the considered objects (images) can be referred to one and only class from several ones. In the case under consideration, the classes are the colour groups and categories of goods. The main stages of building the model for identifying the colour of goods from the fly page: "Uploading the marked images", "Preparing the data for building the model", "Building the model and "Training the model". This stage of uploading assumes unpacking files with the images, uploading the dataset with the image descriptions including the colour, and testing the data integrity. The stage of preparing the data for building the model includes: the formation of the class list for training the model; formation of a file structure for sorting the images; distribution of the images into files. Along with the distribution of the images of the same colour into catalogues, training, validating and test sets are formed according to the requirements of the library for building and training of classification models (Keras). The construction of the model is the crucial stage of its creation. It is at this stage when the model tuning parameters are determined which influence its efficiency. The stage of training the model includes the compilation of the model, initialization of a "generator" of training and testing images and training itself. The compilation of the model implies the call of the function with the same name for the constructed model and estimation of the parameters: the functions of the loss calculation, optimizer and metrics.

The loss calculation function is also called the target function, since it determines how the trained neural network will estimate how close the obtained result is to the expected one. The loss function receives the prediction given by the network and the true value (which the network had to return) and calculates an estimate of the distance between them, reflecting how well the network handled this particular example. The efficiency of neural networks lies in the application of a target function to tune the values of the weights of neurons in order to reduce losses for each image involved in training the model. The tuning itself is performed using an optimizer which implements the so-called error back propagation algorithm, which is the central deep learning algorithm. When training the model, an image generator is used, which is an object responsible for the order in which images are selected from the training and test set

for transferring the model. To train the neural network, it is common practice to recompress images in order to reduce the computational load on the network, the `target_size` parameter determines to what size the generator should reduce the image. Neural network training is performed in batches - subsets of the training set, with the `batch_size` parameter defining the batch size. The classification type is specified by the `class_mode` parameter.

The accumulated database of images and their markup allows one to proceed to the development of algorithms for identifying the colour of goods and their cataloguing for automatically (without the participation of the moderator) dividing newly uploaded items into groups and categories. The first step in solving the problem of image recognition can be considered the construction of convolutional networks and the assessment of their accuracy. Even a very modest result can be interpreted as positive and confirm the possibility of using the proposed approach, subject to further research for estimating the effectiveness of various combinations of the properties of the neural network. Testing of basic configurations will allow determining directions for further research and formulating forthcoming scientific and technical tasks, and obtaining reference values of the classification accuracy to assess the effectiveness of more complex models.

#### **4. Building, training and testing the model for identifying the colour from the image**

The formation of a list of colours consists in analysing the number of images of each colour and discarding categories of colours with low content. To build and train the model for identifying the colour of a product, it was decided to set a threshold value for the number of images of the same colour, namely 80 images. The threshold value allowed us to make a list of 7 popular colours: ['Beige', 'White', 'Blue', 'Brown', 'Pink', 'Gray', 'Black']. The formation of a file structure for sorting images implies the creation of a set of folders with the names corresponding to the target class of the model, i.e. to the colour of the item in the image

To identify the colour of a product from the image, Model 1 was built. It has the sequential structure and consists of 11 layers. The layers are described in Table 1.

The present model is based on convolutional and compression layers, which ensure the identification of individual image features and fixation of these features in the resulting (output) tensor. The number of the detected features coincides with the number of free neurons in the layer. The convolutional layers preserve the size of the input image (taking into account the alignment) and significantly expand the feature space. Compression layers reduce the size of the image in multiples of the subset size, preserving the set of features identified in the previous layer. On layer 8, we get a 7x7 image, where each pixel has 128 features. The ninth (9) layer "unfolds" the three-dimensional tensor into the one-dimensional one, forming a vector of 6272 values at the output. The last two layers are designed to reduce the feature spaces in 2 stages to the required 7 classes corresponding to the colours of the items selected for training.

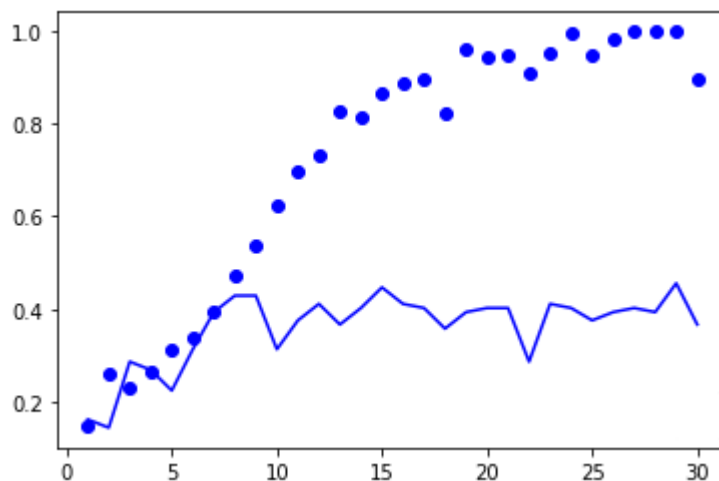
For all but the last layer, the 'relu' activation function was selected. Neurons with this activation function are called ReLU (rectified linear unit). The function has the following formula  $f(x) = \max(0, x)$  and implements a simple threshold transition at zero. On the last layer, the 'softmax' activation function is selected, corresponding to the task of single label multi-class

**Table 1**

The properties of the layers of Model 1

Number of the layer	Type of the layer	Number of neurons	Activation function	Tensor form at the input
1	conv2D, convolutional, core (3, 3)	32	'relu'	(150, 150, 3)
2	MaxPooling2D, convolutional, subset (2, 2)	32	-	(148, 148, 32)
3	conv2D, convolutional, core (3, 3)	64	'relu'	(74, 74, 32)
4	MaxPooling2D, compression, subset (2, 2)	64	-	(72, 72, 64)
5	conv2D, convolutional, core (3, 3)	128	'relu'	(36, 36, 64)
6	MaxPooling2D, compression, subset (2, 2)	128	-	(34, 34, 128)
7	conv2D, convolutional, core (3, 3)	128	'relu'	(17, 17, 128)
8	MaxPooling2D, compression, subset (2, 2)	128	-	(15, 15, 128)
9	Flat	6272	-	(7, 7, 128)
10	Dense	512	'relu'	(6272)
11	Dense	7	'softmax'	(512)

classification. In total, the model under consideration includes 3 456 199 trained parameters. To compile Model 1, we set the loss calculation function - categorical cross entropy, optimizer - RMSProp (root mean square propagation) - a gradient descent algorithm with an impulse, and Accuracy metric which is the portion of correct answers of the algorithm. For Model 1, the image generator of the training set and verification set should produce 150x150 images in batches of 20 images with their category classification. The previously created generator is used as a generator, with the number of steps per epoch being 17, the number of epochs is 30, the number of validation (verification) steps is 6. For 30 epochs of the Model calculation, its accuracy at the training stage was 1, and the value of the loss function was close to 0.

**Figure 1:** The change in the accuracy during the training (.) and the test (-) of Model 1

**Table 2**  
The configuration of Model 2

Number of the layer	Type of the layer	Number of neurons	Activation function	Tensor form at the input
1	conv2D, convolutional, core (3, 3)	32	'relu'	(150, 150, 3)
2	MaxPooling2D, convolutional, subset (2, 2)	32	-	(148, 148, 32)
3	conv2D, convolutional, core (3, 3)	64	'relu'	(74, 74, 32)
4	MaxPooling2D, compression, subset (2, 2)	64	-	(72, 72, 64)
5	conv2D, convolutional, core (3, 3)	128	'relu'	(36, 36, 64)
6	MaxPooling2D, compression, subset (2, 2)	128	-	(34, 34, 128)
7	Flat	369928	-	(17, 17, 128)
8	Dropout, 0,5	-	-	-
9	Dense	512	'relu'	(36992)
10	Dense	6	'softmax'	(512)

The graph of the change in accuracy at the training stage and at the test stage (Figure 1) indicates that after the 8th epoch of calculation, an overtraining effect appeared and at the same stage an acceptable level of accuracy had already been achieved, which was 0.4286, with the value of the loss parameter at the training stage being 0.8321. For the algorithm identifying the colour of the item from the image, the accuracy of the random classification was 0.1339.

## 5. Building, training and testing of the model for determining the clothes category of the item from the image

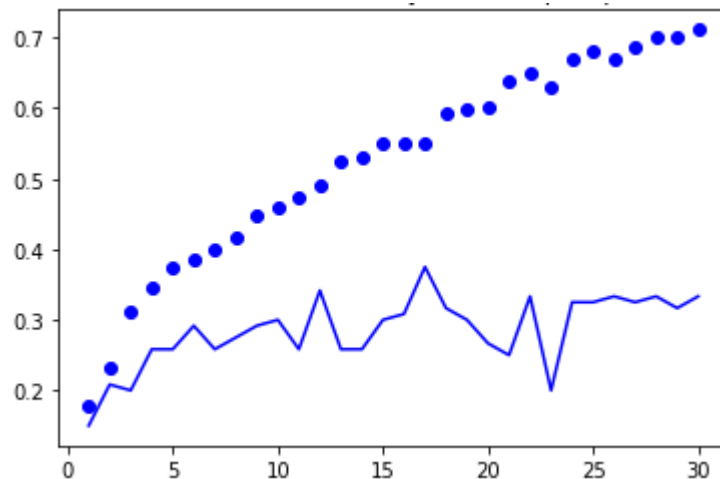
The problem of developing an algorithm for the distribution and cataloguing of goods, similarly to the problem of identifying the colour, can be formulated as a problem of single label multiclass classification.

After unpacking, uploading and checking the data integrity, the analysis of the degree of filling of certain categories with image files was performed. 6 categories of clothing were identified, containing more than 100 photographs: ['Outerwear', 'Suits, outfit', 'Dresses', 'Sweatshirts, Sweaters, Jumpers', 'Bags', 'Decorations']. To train the neural network, it is important that the classes be balanced, i.e. so that each category should contain approximately the same number of images, and there is enough data to fully "configure" the network. In our case, quite few images were accumulated, allowing us only to investigate the fundamental possibility of applying the proposed approach, but not to obtain a ready-made solution with a sufficient level of accuracy.

For the distribution and cataloguing of goods, Model 2 was built. Model 2 has a sequential structure and consists of 10 layers. The layers are described in Table 2.

A Dropout layer is added to Model 2, which zeroes out some revealed features (50% in our case), which allows avoiding the effect of overtraining the neural network. In total, the considered model includes 19,036,742 training parameters.

For Model 2, an infinite training set image generator is created. The generator can generate an almost infinite number of images based on the original set by changing the parameters:



**Figure 2:** The change in the accuracy during the training (.) and the test (-) of Model 2

scale (rescale), image rotation (rotation\_range), width shift (width\_shift\_range), height shift (height\_shift\_range), counterclockwise pixel shift (shear\_range), scaling (zoom\_range), horizontal mirroring (horizontal\_flip). The parameter values set the range in which the generator chooses random values for a particular distortion and applies them to real images. Thus, we expand the training set to infinity.

For Model 2, the following parameter values are defined: rotation\_range = 40, width\_shift\_range = .2, height\_shift\_range = .2, shear\_range = .2, zoom\_range = .2, horizontal\_flip = True. To train Model 2, the previously created generator is used as a generator, the number of steps per epoch is 100, the number of epochs is 30, and the number of validation (verification) steps is 6. For 30 epochs of the Model 2 calculation, its accuracy at the training stage was equal to 0.7115, and the value of the loss function was 0.8184. The graphs of changes in the accuracy at the training stage and at the verification stage (Figure 4) diverge at the 5th epoch of the neural network training. At the 4th step of training, the level of accuracy at the verification stage was 0.2583, with the value of the loss parameter at the training stage being 1.6360. The maximum local accuracy at the verification stage was equal to 0.3750 at epoch 17, but the losses at the training stage at this stage reached 3.0128.

Even small values of the model accuracy parameter at the verification stage can be interpreted as success if they exceed the accuracy parameters of the so-called "random model". Using the random number generator, one of the target classes is randomly selected for each image, and then the accuracy of this random prediction is estimated by comparison with the actual classes. For the model of cataloguing images by clothing categories, the accuracy of the random classification was 0.1642.



## 6. Conclusion

The accumulated database of images and their markup will allow moving on to the development of intelligent algorithms for identifying the colour of goods from fly pages and distribution, as well as cataloguing for automatic (without the participation of the moderator) division of newly uploaded goods into groups and categories. To solve the set problems of single label multiclass classification, deep learning neural networks with convolutional layers, a training set generator using distorting the original images and additional training of a previously trained image classification neural network were created and tested. The proposed models show the classification accuracy higher than that of the "random classifier", which can be interpreted as evidence for the consistency of the selected classification tool, provided that the research continues. The results obtained represent a good theoretical and technological groundwork to continue research in the chosen direction. The improvement of the developed models of intelligent classification can be associated with solving such problems as: expanding the actual database of images, searching and testing an ensemble of models that provide preliminary marking or segmentation of images, building hybrid models for analysing both images and metadata and product description text. The development of an information and analytical system for collecting and presenting information about goods from various sources posted on social networks implies expanding the set of product filters and development of additional services to support the search and purchase of goods.

## References

- [1] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.
- [2] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the national academy of sciences* 79 (1982) 2554–2558.
- [3] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain., *Psychological review* 65 (1958) 386.
- [4] J. Lin, P. Sun, J.-R. Chen, L. Wang, H. Kuo, W. Kuo, Applying gray model to predicting trend of textile fashion colors, *The Journal of The Textile Institute* 101 (2010) 360–368.
- [5] A. Alamsyah, M. A. A. Saputra, R. A. Masrury, Object detection using convolutional neural network to identify popular fashion product, in: *Journal of Physics: Conference Series*, volume 1192, IOP Publishing, 2019, p. 012040.
- [6] K. Hara, V. Jagadeesh, R. Piramuthu, Fashion apparel detection: the role of deep convolutional neural network and pose-dependent priors, in: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2016, pp. 1–9.
- [7] K. Meshkini, J. Platos, H. Ghassemian, An analysis of convolutional neural network for fashion images classification (fashion-mnist), in: *International Conference on Intelligent Information Technologies for Industry*, Springer, 2019, pp. 85–95.
- [8] J. Martinsson, O. Mogren, Semantic segmentation of fashion images using feature pyramid networks, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.