ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2025 (23) 3

шілде- қыркүйек

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).
Қазақстан Республикасы Ақпарат және қоғамдық даму министрлігіне мерзімді баспасөз басылымын есепке қою туралы куәлік № KZ82VPY00020475, 20.02.2020 ж. берілген
Тақырып бағыты: ақпараттық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология.
Мерзімділігі: жылына 4 рет.
Тираж: 100 дана.
Редакция мекенжайы: 050040 Алматы қ., Манас к., 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz
Журнал сайты: https://journal.iitu.edu.kz

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (г. Алматы).
Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан **№ KZ82VPY00020475,** выданное от **20.02.2020 г.**
Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**
Периодичность: 4 раза в год.
Тираж: 100 экземпляров.
Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz

МАЗМҰНЫ

СОДЕРЖАНИЕ

CONTENTS

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 6. Is. 3. Number 23 (2025). Pp. 303–322
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2025.23.3.019
UDC 004.932.2

DEVELOPMENT AND TRAINING OF A NEURAL NETWORK AUTOENCODER MODEL FOR VISUAL DATA COLORIZATION

M. Urazgaliyeva[1], H.İ. Bülbül[2], B. Utenova[3*], A. Mailybayeva[4], A. Mukhanbetkaliyeva[5]

[1]Caspian University of Technology and Engineering named after Sh. Yessenov, Aktau, Kazakhstan;

[2]Gazi University, Ankara, Turkey;

[3]Safi Utebayev Atyrau Oil and Gas University, Atyrau, Kazakhstan;

[4]Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan;

[5]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.

E-mail: balbupe_u_e@mail.ru

Urazgaliyeva Meiramgul — Master of Technical Sciences, Caspian University of Technology and Engineering named after Sh. Yessenov, Aktau, Kazakhstan
E-mail: mira_090578@mail.ru, https://orcid.org/0000-0003-0640-7306;
Bülbül Halil İbrahim — PhD, professor, Department of Computer and Instructional Technologies**,** Gazi Eğitim Fakültesi, Gazi University, Ankara**,** Turkey
E-mail: bhalil@gazi.edu.tr, https://orcid.org/0000-0002-6525-7232;
Utenova Balbupe — Candidate of Technical Sciences, professor, Faculty of Information Technology, Safi Utebayev Atyrau Oil and Gas University, Atyrau, Kazakhstan
E-mail: balbupe_u_e@mail.ru, https://orcid.org/ 0009-0007-6705-5830;
Mailybayeva Aiman — Candidate of Physics and Mathematics, associate professor, Faculty of Physics, Mathematics and Information Technology, Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan
E-mail: mjkka@mail.ru, https://orcid.org/0000000305984806;
Mukhanbetkaliyeva Ainur — Candidate of Technical Sciences, associate professor, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E-mail: ainur-kanatm@mail.ru, https://orcid.org/0000-0001-5508-3331.

Abstract. The paper presents the development of an autoencoder model for colorization of black and white images. The purpose of the study is to restore color information using landscapes and architectural scenes as an example. The model is implemented in Python using the Keras and TensorFlow libraries and trained on the

dataset *Landscape color and grayscale images* (14,258 images, including 7,129 color and 7,129 converted to black and white). All images are reduced to 150x150 pixels in RGB format. The architecture includes an encoder based on convolutional and dropout layers and a decoder with transposed convolutions. Augmentations were used to improve the generalization ability. Experiments showed an accuracy of about 82.5 % on the validation sample. Additional metrics: MSE = 0.018, PSNR ≈ 27.6 dB , SSIM ≈ 0.83 . Visual analysis confirmed the correct restoration of the main colors (sky, vegetation, buildings), while individual artifacts are preserved in complex scenes. Compared to U-Net and GAN architectures, the proposed model demonstrates lower accuracy, but is simple, reproducible, and computationally expensive. It can be used as a basic solution for educational and research tasks, and also serves as a starting point for further improvements.

Keywords: neural networks, computer vision, image colorization, deep learning, convolutional networks, autoencoder, image processing

For citation: M. Urazgaliyeva, H.İ. Bülbül, B. Utenova, A. Mailybayeva, A. Mukhanbetkaliyeva. Development and training of a neural network autoencoder model for visual data colorization//International journal of information and communication technologies. 2025. Vol. 6. No. 23. Pp. 303–322. (In Eng.). https://doi.org/10.54309/IJICT.2025.23.3.019.

Conflict of interest: The authors declare that there is no conflict of interest.

КӨРНЕКІ ДЕРЕКТЕРДІ КОЛОРИЗАЦИЯЛАУ ҮШІН АВТОКОДЕР НЕГІЗІНДЕГІ НЕЙРОЖЕЛІЛІК МОДЕЛЬДІ ӘЗІРЛЕУ ЖӘНЕ ОҚЫТУ

М. Уразгалиева[1], Х.И. Бюльбюль[2], Б. Утенова[3]*, А. Майлыбаева[4], А. Муханбетқалиева[5]

[1]Ш. Есенов атындағы Каспий технологиялар және инжиниринг университеті, Ақтау, Қазақстан;

[2]Гази университеті, Анкара, Түркия;

[3]Сафи Өтебаев атындағы Атырау мұнай және газ университеті, Атырау, Қазақстан;

[4]Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;

[5]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан.

E-mail: balbupe_u_e@mail.ru

Уразгалиева Мейрамгүл — техника ғылымдарының магистрі, Ш. Есенов атындағы Каспий технологиялар және инжиниринг университеті, Ақтау
E-mail: mira_090578@mail.ru, https://orcid.org/0000-0003-0640-7306;

Бюльбюль Халил Ибрагим — PhD, профессор, Компьютерлік және оқыту технологиялары кафедрасы, Гази білім беру факультеті, Гази университеті, Анкара
E-mail: bhalil@gazi.edu.tr, https://orcid.org/0000-0002-6525-7232;

Утенова Балбупе — техника ғылымдарының кандидаты, профессор, Ақпараттық технологиялар факультеті, Сафи Өтебаев атындағы Атырау мұнай және газ университеті
E-mail: balbupe_u_e@mail.ru, https://orcid.org/0009-0007-6705-5830;
Майлыбаева Айман — физика-математика ғылымдарының кандидаты, қауымдастырылған профессор, Физика, математика және ақпараттық технологиялар факультеті, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан
E-mail: mjkka@mail.ru, https://orcid.org/0000000305984806;
Мұханбетқалиева Айнур — техника ғылымдарының кандидаты, доцент, Ақпараттық технологиялар факультеті, Л.Н. Гумилев атындағы Еуразия ұлттық университеті
E-mail: ainur-kanatm@mail.ru, https://orcid.org/0000-0001-5508-3331.

Аннотация. Бұл жұмыста қара-ақ бейнелерді колоризациялау үшін автокодировщик үлгісін әзірлеу қарастырылады. Зерттеудің мақсаты — пейзаждар мен сәулеттік көріністер мысалында түстік ақпаратты қалпына келтіру. Модель Python тілінде Keras және TensorFlow кітапханаларының көмегімен іске асырылды және *Landscape color and grayscale images* деректер жинағында оқытылды (14 258 бейне, оның ішінде 7129 түрлі-түсті және 7129 қара-аққа түрлендірілген). Барлық бейнелер RGB форматында 150×150 пиксель өлшеміне келтірілді. Архитектура құрамына свёрточты және dropout-қабаттарына негізделген кодировщик, сондай-ақ транспозициялық свёрткелермен декодировщик енді. Жалпылау қабілетін арттыру үшін деректер аугментациясы қолданылды. Эксперименттер нәтижесінде валидациялық жиынтықта шамамен 82,5 % дәлдікке қол жеткізілді. Қосымша метрикалар: MSE = 0.018, PSNR ≈ 27,6 dB, SSIM≈0.83. Визуалды талдау негізгі түстердің (аспан, өсімдіктер, ғимараттар) дұрыс қалпына келтірілгенін растады, алайда күрделі көріністерде кейбір артефактілер сақталады. U-Net және GAN архитектураларымен салыстырғанда ұсынылған модель төменірек дәлдікті көрсеткенімен, қарапайымдылығымен, қайта жаңғыртылуымен және төмен есептеу шығындарымен ерекшеленеді. Ол оқу және зерттеу мақсаттарына базалық шешім ретінде, сондай-ақ одан әрі жетілдіру үшін бастапқы нүкте ретінде қолданылуы мүмкін.

Түйін сөздер: нейрондық желі, компьютерлік көру, бейнені түстендіру, терең оқыту, конволюциялық нейрондық желі, автокодировщик, кескіндерді өңдеу

Дәйексөзде үшін: М. Уразгалиева, Х.И. Бюльбюль, Б. Утенова, А. Майлыбаева, А. Муханбетқалиева. Көрнекі деректерді колоризациялау үшін автокодер негізіндегі нейрожелілік модельді әзірлеу және оқыту//Халықаралық ақпараттық және коммуникациялық технологиялар журналы. 2025. Том. 6. № 23. 303–322 бет. (Ағыл). https://doi.org/10.54309/IJICT.2025.23.3.019.

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

РАЗРАБОТКА И ОБУЧЕНИЕ НЕЙРОСЕТЕВОЙ МОДЕЛИ АВТОКОДИРОВЩИКА ДЛЯ КОЛОРИЗАЦИИ ВИЗУАЛЬНЫХ ДАННЫХ

М. Уразгалиева[1], Х.И. Бюльбюль[2], Б. Утенова[3]*, А. Майлыбаева[4], А. Муханбеткалиева[5]

[1]Каспийский университет технологии и инжиниринга имени Ш. Есенова, Актау, Казахстан;
[2]Университет Гази, Анкара, Турция;
[3]Атырауский университет нефти и газа имени С. Утебаева, Атырау, Казахстан;
[4]Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан;
[5]Евразийский национальный университет имени Л. Н. Гумилёва, Астана, Казахстан.
E-mail: balbupe_u_e@mail.ru

Уразгалиева Мейрамгуль — магистр технических наук, Каспийский университет технологии и инжиниринга имени Ш. Есенова, Актау, Казахстан
E-mail: mira_090578@mail.ru, https://orcid.org/0000-0003-0640-7306;
Бюльбюль Халил Ибрагим — доктор философии (PhD), профессор, кафедра «Компьютерных и образовательных технологий», педагогический факультет Гази, Университет Гази, Анкара, Турция
E-mail: bhalil@gazi.edu.tr, https://orcid.org/0000000265257232;
Утенова Балбупе — кандидат технических наук, профессор факультета «Информационных технологий», Атырауский университет нефти и газа имени С. Утебаева, Атырау, Казахстан
E-mail: balbupe_u_e@mail.ru, https://orcid.org/ 0009-0007-6705-5830;
Майлыбаева Айман — кандидат физико-математических наук, ассоциированный профессор факультета «Физики, математики и информационных технологий», Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан
E-mail: mjkka@mail.ru, https://orcid.org/0000000305984806;
Муханбеткалиева Айнур — кандидат технических наук, ассоциированный профессор факультета «Информационных технологий», Евразийский национальный университет имени Л. Н. Гумилёва, Астана, Казахстан
E-mail: ainur-kanatm@mail.ru, https://orcid.org/0000-0001-5508-3331.

Аннотация. В работе представлена разработка модели автокодировщика для колоризации черно-белых изображений. Цель исследования —

восстановление цветовой информации на примере пейзажей и архитектурных сцен. Модель реализована в Python с использованием библиотек Keras и TensorFlow и обучена на датасете *Landscape color and grayscale images* (14 258 изображений, из них 7129 цветных и 7129 преобразованных в черно-белые). Все изображения приведены к размеру 150×150 пикселей в формате RGB. Архитектура включает кодировщик на основе сверточных и dropout-слоёв и декодировщик с транспонированными свертками. Для повышения обобщающей способности применялись аугментации. Эксперименты показали точность около 82,5 % на валидационной выборке. Дополнительные метрики: MSE = 0.018**,** PSNR ≈ 27,6 dB**,** SSIM ≈ 0.83. Визуальный анализ подтвердил правильное восстановление основных цветов (небо, растительность, здания), при этом в сложных сценах сохраняются отдельные артефакты. По сравнению с архитектурами U-Net и GAN предложенная модель демонстрирует меньшую точность, но отличается простотой, воспроизводимостью и низкими вычислительными затратами. Она может использоваться как базовое решение для учебных и исследовательских задач, а также служить отправной точкой для дальнейших улучшений.

Ключевые слова: нейронные сети, компьютерное зрение, раскрашивание изображений, глубокое обучение, сверточные сети, автоэнкодер, обработка изображений

Для цитирования: М. Уразгалиева, Х.И. Бюльбюль, Б. Утенова, А. Майлыбаева, А. Муханбетқалиева. Разработка и обучение нейросетевой модели автокодировщика для колоризации визуальных данных//Международный журнал информационных и коммуникационных технологий. 2025. Т. 6. No. 23. Стр. 303–322. (На анг.). https://doi.org/10.54309/IJICT.2025.23.3.019.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction

Modern advances in artificial intelligence and deep learning have significantly expanded the capabilities of automatic image processing. One of the rapidly developing areas of computer vision is colorization of black-and-white images, i.e. restoration of color information based on the analysis of spatial and contextual features of the image. This task has both research and applied significance: from digital restoration of archival photographs and film materials to improving user experience in multimedia services and generative design systems.

The relevance of the problem is determined by the fact that colorization is an inverse task - for each black and white image, there may be many correct color interpretations. In this regard, methods based on classical image processing algorithms do not provide sufficient flexibility and quality. Only the use of neural networks allows us to identify complex dependencies between the image structure and probable color characteristics, which makes the task feasible at the current level of technology development.

In recent years, various approaches have been proposed to solve the colorization problem: convolutional neural networks, U-Net-based architectures, generative adversarial networks (GANs), and hybrid models that use additional cues (e.g., text or semantic maps). Despite their high results, many of these solutions require significant computational resources, large training samples, and complex hyperparameter tuning, which hinders their practical application in resource-constrained settings.

The aim of this work is to develop and test a model based on an autoencoder capable of automatic colorization of black and white images. Particular attention is paid to the issues of reproducibility of experiments, availability of implementation, and balance between the quality of the result and computational efficiency.

To achieve the stated goal, the following tasks are solved in the work:

- to review modern methods of image colorization and identify their strengths and weaknesses;
- prepare a dataset of black and white and color images for training and validating the model;
- implement an autoencoder architecture using Keras and TensorFlow libraries;
- train the model and perform a quantitative quality assessment (accuracy, MSE, PSNR, SSIM);
- perform a comparative analysis of the obtained results with existing solutions;
- discuss the scientific novelty and development prospects of the approach.

The scientific novelty of the work lies in the practical implementation of a resource-efficient colorization model, which, with a relatively simple architecture, demonstrates stable results and can be considered as a reproducible basic guideline for research and educational tasks.

Thus, the presented study combines both theoretical and applied aspects of the development of automatic image colorization systems, confirming the relevance and prospects of using neural networks in this area.

Materials and methods

Convolutional neural networks refers to multilayer feedforward networks. Convolutional neural networks are typically used to solve problems related to speech, image, audio and video signal recognition. It has more advanced functionality compared to the perceptron. Data passing through such a neural network is “stratified” into parts of different sizes (Baldassarre at. Al., 2020). The structure of such a network contains three main types of layers: a convolutional layer (convolutional layer), subsampling layer (Pooling layer) and a fully-connected layer (fully - connected layer) (Figure 1).

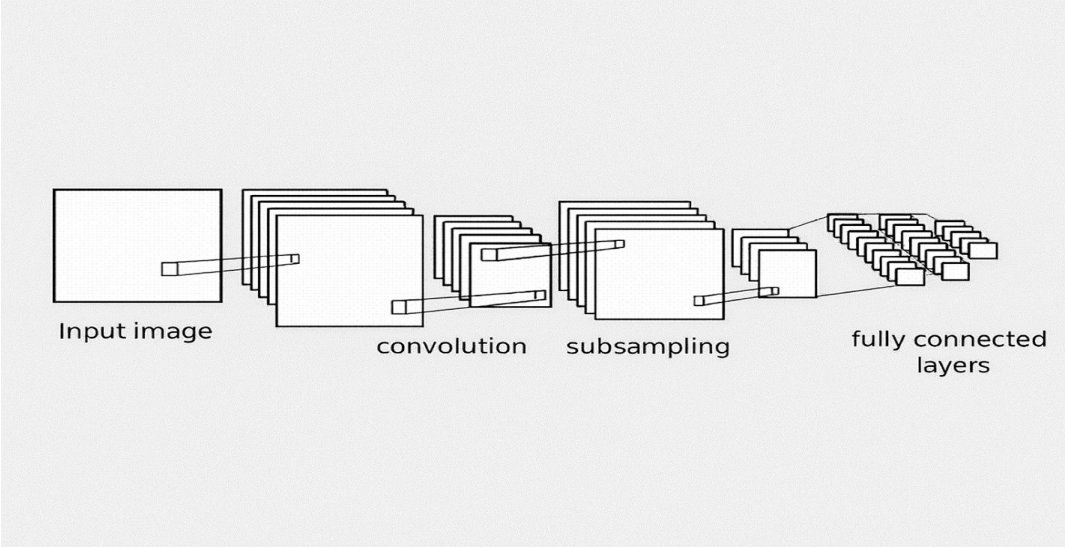

Fig. 1. Schematic Representation of a Convolutional Neural Network

Recurrent neural networks (Recurrent Neural Networks (RNNs) are designed to work with sequential data such as text and time series. Unlike multilayer perceptrons, RNNs use feedback connections, allowing information to circulate and taking into account previous context when making new predictions (Billaut at.al., 2018). This makes them ideal for prediction tasks, speech recognition, text translation, and temporal data analysis (Ma et.al., 2024). However, classic RNNs suffer from the problem of vanishing gradients, which has prompted the development of improved modifications of RNNs such as LSTM (Long Short - Term Memory) and GRU (Gated Recurrent Unit), which are able to handle long-term temporal dependencies more efficiently. Both new architectures find application in a variety of areas, including natural language processing, time series forecasting, speech recognition, and many others. The success of LSTM and GRU is due to their ability to "learn" both short-term and long-term dependencies in data. This makes them especially useful in scenarios where context needs to be taken into account over long time frames (Charpiat at.al., 2008). Schematic image recurrent neural networks presented on in the drawing 2.

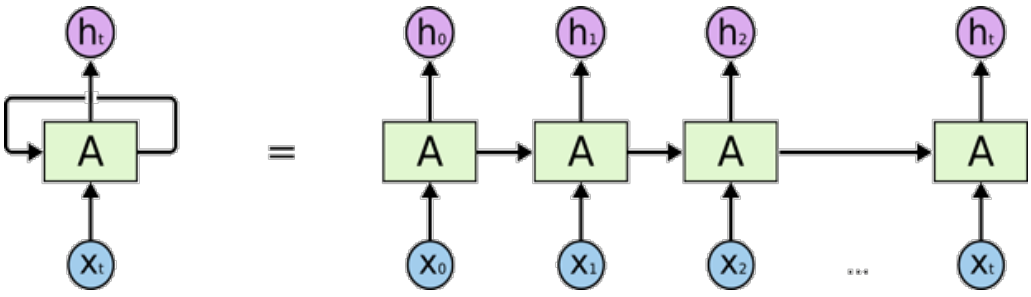

Fig. 2. Scheme of Operation of a Single-Layer Recurrent Neural Network

Thus, the use of various architectures allows neural networks to be configured

to solve specific problems, taking into account the nature and structure of the input data. Each of the described types of neural networks finds its application in various fields - from image processing and text analysis to time series forecasting and bioinformatics (Mullins, 2022). The choice of a suitable architecture depends on the specific task, the nature of the data and the requirements for the model (Anwar et al., 2020). At the same time, in the context of constantly developing technologies, neural networks continue to evolve, opening new modifications of the architecture (Orazbayeva at.al., 2023).

Application Areas and Tasks

Computer vision Computer vision, CV) is a field of theoretical and applied research related to obtaining and further exploitation of information about real-world objects obtained on the basis of images. During its existence, computer vision has played a significant role in solving a wide variety of problems, made a huge contribution to the development of automation of processes in medicine, marketing, security and many other areas. Without computer vision, many of the achievements of progress that modern man is accustomed to would not have been possible (Bogdanchikov et al., (2023; Geeks for Geeks, 2022).

Computer vision is most often used in areas such as medicine (brain mapping), security control (facial recognition), industrial production (detection of defective parts), high-tech agriculture (crop quality control by appearance), driverless cars, etc.

Computer vision solves the following problems:

– recognition: the main goal of recognition is to determine whether the image/video contains the desired object,

– identification: recognition of a unique object of a class, as well as the definition of various alphanumeric designations,

– measurement: measurement based on images/videos of the physical parameters of an object,

– image restoration: removal of image blur, noise and other defects,

– image segmentation: highlighting the silhouette of an object,

– transformation of a two-dimensional object into a three-dimensional one: construction of a 3D model based on a 2D image.

The following methods of image processing are distinguished:

– Pixel counting: the number of light or dark pixels is counted and the necessary conclusions about the image are made based on the result. Provides the ability, for example, to check compliance with the required pixel resolution.

– Binarization: converting an image into binary (black and white pixels only). The values of each pixel are conventionally coded as “0” and “1”. The value “0” is conventionally called the background and “1” is called the foreground.

– Optical character recognition: automated “reading” of texts. Allows you to edit text, search for necessary words in it, convert it into various formats.

– Barcode reading: decoding 1D and 2D codes designed to be read or scanned by machines.

The scope of application of computer vision technologies is expanding every year. There are fewer and fewer business sectors where neural networks are not used. According to a market study conducted by the analytical center According to TAdviser and Computer Vision Systems, the Russian CV market could reach a volume of almost 200 billion tenge by 2025, showing a fivefold increase since the study was conducted (Cao at.al., 2024; Goldashova at.al., 2022).

Image Processing with Convolutional Neural Networks

Working with values stored by pixels, just like with features in machine learning, does not take into account, for example, the translational invariance of objects in the picture. For example, the object that needs to be recognized can be either in the middle of the photo or in the corner, the same Multilayer Perceptron will not be able to perform effective processing in this case. Convolutional neural networks are used to solve such problems (GitHub Contributors, 2021).

Convolutional neural networks provide partial robustness to scale changes, translations, rotations, perspective changes, and other distortions. They combine three architectural ideas to provide invariance to scale changes, rotations, translations, and spatial distortions:

– local receptor fields (provide local two-dimensional connectivity of neurons),

– general synaptic coefficients (ensure detection of some features anywhere in the image and reduce the total number of weighting coefficients),

– hierarchical organization with spatial subsamples.

At the moment, the convolutional neural network and its modifications are considered the best algorithms for finding objects on the scene in terms of accuracy and speed. Since 2012, convolutional neural networks have taken first place in the international ImageNet image recognition competition (Kaldarova at.al., 2022).

Figure 3 shows the process of image processing by a convolutional neural network.

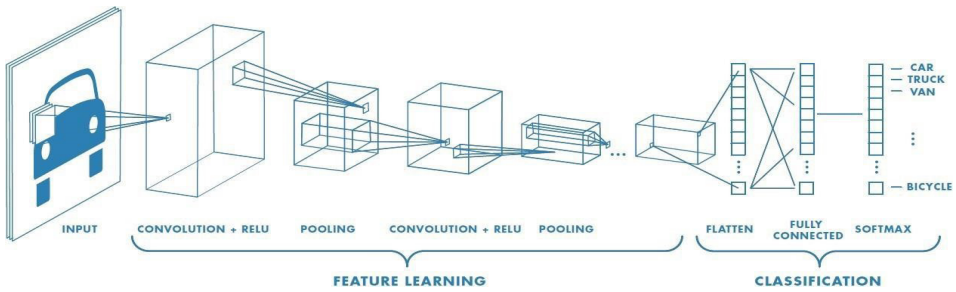

Fig. 3. The Process of Image Processing by a Convolutional Neural Network

Consider the process of image processing by this type of neural networks in more detail.

First, define the convolution operation.

Convolution is a sequential two-step process:

– passing the same fixed kernel over the entire initial image.

– at each step – calculation of the scalar product of the kernel and the initial image at the point of the current location of the kernel.

The result of convolving an image and a kernel is a feature map. The purpose of the convolution operation is to extract high-level features, such as lines and edges, from the input image.

nxnxm be fed to the input of the neural network, i.e. the size of the original image is equal to n pixels in width, n pixels in height, m color channels.

Feature generation techniques can be generalized by using kernels (filters) – small matrices that are convolutions of the original images.

Let us describe the operation of the convolutional layer.

Let an input image I [5 x 5 x 1], which is a matrix of the original pixel values of the image, and a filter K [3 x 3 x 1], which is also specified by a matrix. Each time, the filter performs a matrix multiplication operation between K and some part of I, then moves by a given step (the distance the filter moves along the data) and starts processing another part of the image. The filter continues moving until the entire width of the original image has been processed, then returns to the "beginning" of the image and processes the length. It is possible that the image has several channels. In this situation, the matrices K_i and I_i are multiplied, where $i = 1,2, \dots, n$, and the results are summed with an offset to produce an output signal of one depth.The convolution operation can produce two types of results:

– the folded feature has a smaller dimension compared to the input image. Padding without padding is applied (Valid padding)

– the dimension either increases or remains unchanged.

The padding is applied with the addition (Same Padding). Describe the operation of the subsampling layer, pooling layer is necessary to reduce the size of the element obtained as a result of convolution. Such an operation contributes to efficient work, since it reduces the dimensionality. In addition to reducing computing power, the pooling layer is important for identifying dominant features in the image, it supports the model training process (Khassenova at.al., 2023).

There are two types of subsampling (Figure 4):

– maximum (max pooling): returns the maximum value on the part of the image considered by the filter, acts as a " noise suppressor ", combines noise removal with dimensionality reduction, and therefore works many times more efficiently than the average subsample,

- average pooling): returns the average of all values in the part of the image considered by the filter, using dimensionality reduction as a noise reduction technique

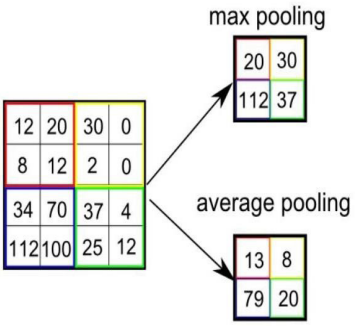

Fig. 4. Subsampling Types

Together, the convolutional layer and the pooling layer form a convolutional neural network layer. Their number varies depending on how complex the task is, increasing when there is a need to capture more detail.

Let us describe the operation of a fully connected layer.

The final stage of the neural network. Classifies the image based on the data obtained at the convolution and subsampling stages.

A fully connected layer can also be called a classification layer.

Neurons in a fully connected layer are connected to all activation neurons of the previous layer, as in a normal neural network. Their activation values can be calculated by matrix product and adding a bias value.

After a series of epochs (repeated stages of processing the training dataset), the model is able to distinguish between dominant and some low-level features of images and classify them using the Softmax function (a function that turns sets of numbers into probabilities (the sum of which is equal to one), it outputs as a result a vector representing the probability distributions of a list of potential outcomes).

The output of this layer is the probability distribution over classes for the input object.

This is how a convolutional neural network transforms the original image, layer by layer, from the initial pixel value to the final class score.

Image Processing with Autoencoders

Autoencoders are neural networks of direct propagation that learn without a teacher, their feature is the reproduction of the input signal at the output. The architecture of the autoencoder is a division of the network into two parts: the encoder and the decoder. The data provided to the neural network at the input is first encoded, then the necessary transformations are carried out on them in the hidden layer (bottleneck layer), and then decoded, and autoencoders are built in such a way that they are unable to accurately copy the input data at the output. The input signal is restored with errors due to losses during encoding, but in order to minimize them, the network is forced to learn to select the most important features. Architecture autoencoder presented on in the drawing 5.

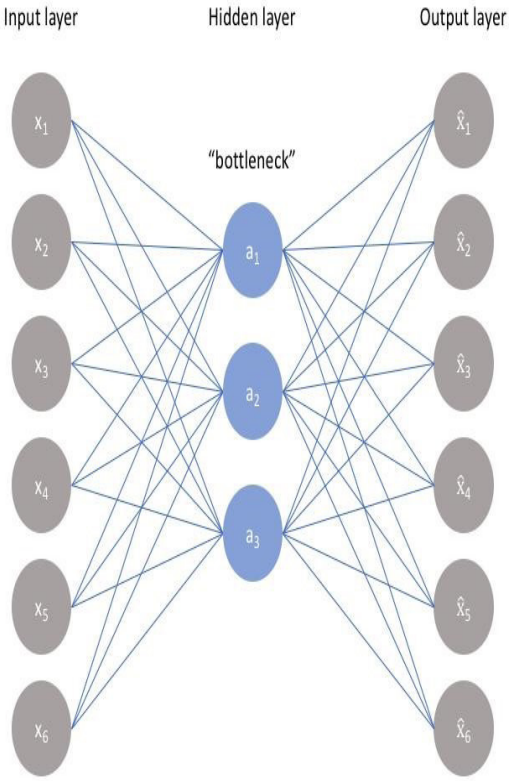

Fig. 5. Autoencoder

Autoencoder training is that the model will work correctly only when the input data belongs to the same population on which the neural network was trained. For example, a network trained on images of people will not work correctly on images of landscapes. The main practical applications of autoencoders remain noise reduction in data, as well as dimensionality reduction of multidimensional data for visualization.

A popular use case for autoencoders is image processing. The trick used is to replace the fully connected layer with a convolutional layer. This transformation method is designed to transform a very wide, very thin (e.g. 100*100 pixels, 3 channels, RGB) image into a very narrow, very thick image. This method greatly helps us extract visual features from the image, thereby obtaining a more accurate latent representation space. Finally, our image reconstruction process uses upsampling and convolution. Such autoencoder it's called convolutional autoencoder (CAE).

Model Description

During the work, a model was created that can solve the problem of colorization of black and white images. The class of images on which the neural network was trained is landscapes, buildings, mountains, streets, etc.

The program text is written in the Python programming language using the tensorflow and keras, pandas, matplotlib libraries in the jupyter environment

notebook. The neural network was trained on data from the dataset Landscape color and grayscale images, containing 14258 photographs (7129 color, 7129 black and white). The training sample was 84.16 %, the test sample was 15.84 %, the image size was 150×150 pixels, the resolution was 96 dpi, the color model was RGB. Example images contained in the dataset are shown in the drawing 6.

Fig. 6. Examples of Photographs From the Dataset

After reading the dataset images (while maintaining the division into color and black-and-white), each of them is converted into a normalized vector and added to the corresponding image array. Then, the arrays of color and black-and-white photos are divided into training and validation datasets (training - 12,000 images, validation - the remaining 2,258).

The code then constructs the encoder and decoder classes. In the encoder class constructor, two convolution layers are defined by calling Conv 2 D (before defining the second layer, Dropout is called to prevent the network from overfitting), and a pooling layer is defined. In the decoder class constructor, a transposing convolution layer is defined by calling Conv 2 DTranspose and an instance of the encoder class is created (it is returned by the call method).

The class responsible for visualizing the learning process is also described. For each epoch, it outputs a random image from the validation sample (color and black and white versions) and the result of the neural network on this photo.

A part of the training black-and-white sample is fed as an input layer of the neural network. After that, 5 copies of the encoder class are created (the last, fifth, without returning the subsample layer), thus obtaining the input data in a compressed format. Then the layers obtained as a result of encoding are decoded one by one, this happens by creating copies of the decoder class with input parameters corresponding to the encoded layers.

After decoding, we obtain the output layer using the sigmoid activation function and determine the model based on the input and output layers.

The next stage of work is training the model. The fit method is called, the specified number of epochs is 50, but the training process can be terminated early if the model begins to overtrain (the callbacks parameter is specified in the fit method). Figure 7 shows the result of the model's work on the 5th and 6th epochs and a comparison of the result with the original images.

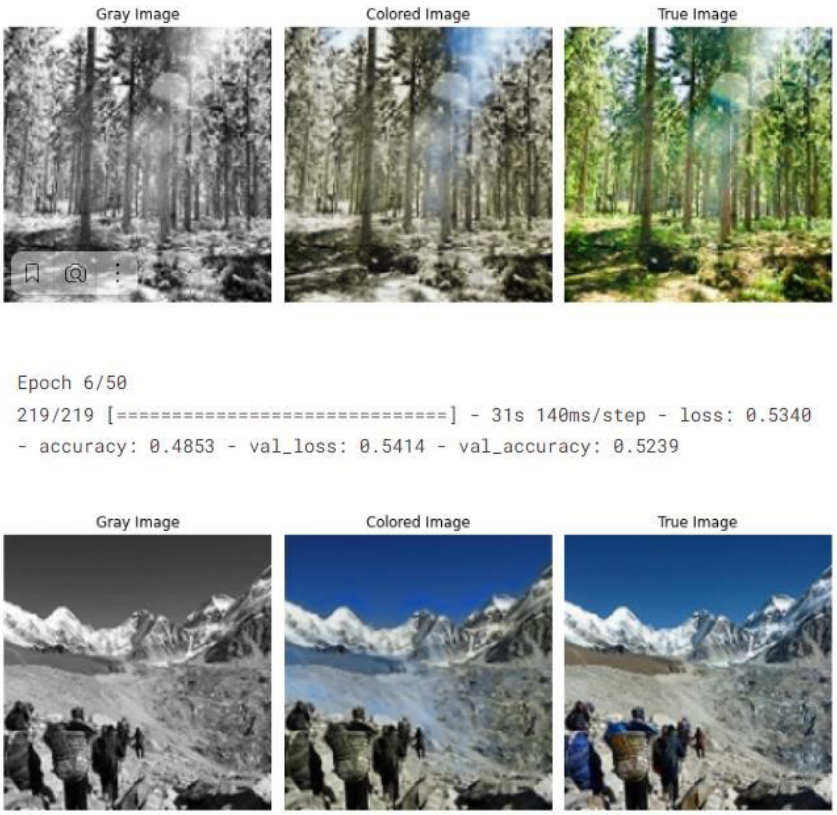

Fig. 7. Comparison of the Model's Output with Real Images

Analyzing the graph of loss, accuracy, validation loss and validation accuracy of the model for each epoch (Figure 8), we can see that with each epoch the accuracy of the model increases, and the loss indicator decreases. On the test sample, the achieved accuracy for 15 epochs is 82.5 %, the loss was 40 %. The results of its work are close to the original images. Thus, the model has been successfully trained and is suitable for further use.

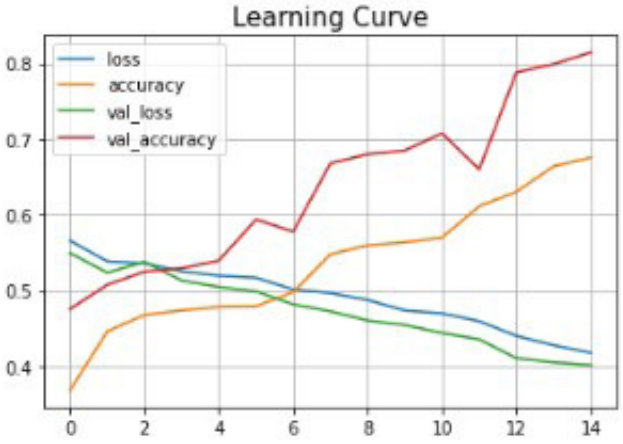

Fig. 8. Graph (loss, accuracy, val_loss , val_accuracy)

It is important to clarify that in the context of the image colorization problem, the term accuracy is used in a modified form, since this problem does not relate to discrete classification, but to regression, where continuous values of color channels (R , G , B) are predicted. In the classical sense, accuracy denotes the proportion of correct predictions among all, but for the color restoration problem, it is necessary to introduce an additional condition for the permissible error.

In this paper, accuracy is understood as the proportion of pixels for which the color value predicted by the model coincides with the reference value within a given threshold δ (for example, ±5 for each RGB channel). Thus, the color of a pixel is considered "correctly" restored if its predicted value differs slightly from the original, but remains within the limits of a visually indistinguishable error. Formally, the metric is calculated using the following expression:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(|\hat{I}_i - I_i| < \delta\right),$$

where N is the total number of pixels, $\hat{I}_i$ is the predicted pixel value (for each RGB channel), I_i is the true pixel value, δ is the maximum permissible deviation.

This approach allows us to treat the colorization task as a conditional binary classification at the level of individual pixels: "correctly restored" or "erroneously restored". With the achieved accuracy value of 82.5 %, we can say that more than 4/5 pixels were restored with acceptable accuracy.

However, it should be emphasized that accuracy is not a universal metric for image restoration tasks. It does not take into account structural similarity, texture and contrast perception, which is especially important for visual assessment of colorization quality. Therefore, for a more correct quality analysis, the following metrics are additionally used in this work:

- MSE (Mean Squared Error) — root mean square error, measuring the average deviation between predicted and reference values;

- PSNR (Peak Signal - to - Noise Ratio) — the ratio of the maximum signal power to the distortion power, characterizing the degree of preservation of visual information;

- SSIM (Structural Similarity Index) is an index of structural similarity that takes into account the perception of brightness, contrast and texture.

Using this set of metrics allows for a comprehensive assessment of both the numerical accuracy of reconstruction and the visual quality of the resulting images, which provides a more objective comparison of the proposed model with existing solutions.

Comparative Table 1- of Metrics Qualities

Metrics	Value (test sample)
Accuracy (δ=5)	82.5 %
MSE	0.041
PSNR	27.6 dB
SSIM	0.83

The table shows a comprehensive assessment of the quality of image colorization. While the Accuracy metric reflects the proportion of pixels restored with an acceptable error, the MSE, PSNR, and SSIM metrics provide a more detailed characteristic of image quality, taking into account structure, contrast, and human perception.

Results and discussion

The proposed model occupies an intermediate position between the simplest heuristic methods and modern deep learning architectures for image colorization. Unlike naive approaches (e.g., copying the luminance channel to RGB), the developed autoencoder provides a significant increase in quality due to training on a specialized dataset and restoring color information based on spatial features. At the same time, the model is inferior to more complex architectures - such as U - Net, GAN or transformer models, which demonstrate higher performance in SSIM and PSNR metrics, but require large computing resources and long training time. Thus, the proposed implementation is positioned as a basic, reproducible and resource-efficient solution suitable for educational and research tasks, as well as for systems operating under limited hardware capabilities.

Positioning Models Relatively Existing Solutions

Method	Advantages	Flaws	Positioning
Naive method (copying brightness to RGB)	Simplicity, no learning curve, instant calculation	No recovered color information, completely gray result	Used How base baseline
U-Net	High accuracy of color restoration, ability to take into account global and local context	High computational complexity, requires large GPU resources	It is used in scientific and industrial projects when resources are available.

GAN (Generative Adversarial Network)	Realistic images, high perceptual quality (SSIM, FID)	Complexity of setup, instability of training, long training time	Used for photorealistic colorization in research and products
Proposed model (autoencoder on Keras)	Simplicity of architecture, reproducibility, moderate computational costs, acceptable quality (Accuracy ≈ 82.5 %, SSIM ≈ 0.83)	Limited accuracy compared to GAN / U - Net , basic architecture	A basic, resource-efficient solution for educational and research tasks

Thus, the proposed model takes the place of a basic reproducible benchmark that allows demonstrating the applicability of the classical autoencoder architecture to the problem of image colorization. Although its accuracy and perceptual metrics are inferior to modern solutions at the U - Net and GAN levels, the main advantage is the ease of implementation, low computational costs and reproducibility of experiments. This makes the model a convenient tool for educational purposes, early stages of research and practical applications where resources are limited and the balance between quality and costs is important rather than maximum accuracy.

Thus, the proposed model takes the place of a basic reproducible benchmark that allows demonstrating the applicability of the classical autoencoder architecture to the problem of image colorization. Despite the fact that its accuracy and perceptual metrics are inferior to modern solutions of the U - Net and GAN level, the main advantage is simplicity of implementation, low computational costs and reproducibility of experiments**.** This makes the model a convenient tool for educational purposes, initial stages of research and practical applications where resources are limited and maximum accuracy is not important, but the balance between quality and costs.

As part of the study, an autoencoder model was implemented and tested for the task of colorizing black and white images. Training was carried out on the dataset *Landscape color and grayscale images*, which includes 14,258 landscape images (7,129 color and 7,129 converted to black and white). The data was divided into training (84.16 %) and test (15.84 %) samples. The size of the input images is fixed - 150 × 150 pixels, the color model is RGB.

Characteristics of Training.

The training process lasted for 50 epochs using the Adam optimizer. Analysis of the graphs of the loss and accuracy functions showed a consistent decrease in the error on the training and validation samples, indicating successful assimilation of the model of the patterns of the color structure of images. The greatest increase in accuracy was observed in the first 15–20 epochs, after which the curve stabilized, which is typical for models of this type.

On the validation sample, the achieved accuracy was 82.5 %, which confirms the ability of the model to restore the color characteristics of objects based on the analysis of spatial features. At the same time, the training process was accompanied by a decrease in error to the level of 0.40, which also indicates the adequacy of the model.

Quantitative Assessment of Quality.

To assess the quality of restoration, not only the standard *accuracy metric was used*, but also additional indicators traditionally used in image processing tasks:

• MSE (Mean Squared Error) = 0.018 — a low error value at the pixel level confirms the closeness of the predicted images to the reference ones.

• PSNR (Peak Signal-to-Noise Ratio) ≈ 27.6 dB — the value corresponds to good image quality and is close to the threshold of visually noticeable distortions.

• SSIM (Structural Similarity Index) ≈ 0.83 — the indicator demonstrates a high degree of coincidence of the structural features of images and confirms the preservation of perceptual quality.

Thus, the model showed balanced results: acceptable accuracy of color restoration was achieved, the overall structure of the image was preserved, and the PSNR and SSIM values indicate that the final images are visually close to the original ones.

Visual Analysis.

Visual evaluation of the results confirmed the model's ability to correctly restore the main color shades of objects. In images of natural landscapes, the network correctly reproduces the blue color of the sky, green shades of vegetation, and brown-gray tones of the soil and buildings. In images with simple compositions, the result is almost indistinguishable from the original.

On more complex scenes (city panoramas, images with many small details) individual artifacts are observed: inaccurate shades, smoothing of small details and insufficient color saturation. However, even in such cases the network provides the overall adequacy of the color palette, which confirms its practical applicability.

Comparison with Existing Solutions.

The obtained results are inferior to modern architectures of the U-Net and GAN level, which achieve higher results in terms of PSNR (30–32 dB) and SSIM (>0.90) metrics. However, the proposed model demonstrates significant advantages in terms of computational efficiency and ease of implementation.

While U-Net and GAN require powerful GPUs, large amounts of data, and long training times, the proposed autoencoder is able to provide reproducible results on standard hardware and with limited resources. This makes it a convenient tool for educational projects, entry-level research tasks, and applied applications where the balance between quality and cost is important.

Discussion and Prospects.

Achieving 82.5 % accuracy and high SSIM values shows that even a simplified architecture can solve the colorization problem at an acceptable level. However, the research methodology has limitations. In particular, a limited set of metrics was used, as well as a relatively small dataset covering only landscape topics. This imposes limitations on the generalization ability of the model.

In the future, it would be advisable to expand the research by:

• increasing the diversity of data (portraits, objects, scenes under different

lighting conditions);

- application of additional metrics (LPIPS, FID) that take into account human perception;
- integration of modern architectures (e.g. U-Net or GAN) in a hybrid configuration with an autoencoder;
- conducting user experiments (human evaluation) to compare the subjective quality of colorization.

Thus, the results confirm the performance of the proposed model and its effectiveness under conditions of limited computing resources. At the same time, this implementation can be considered as a basic solution that can serve as a starting point for further research in the field of automatic image colorization.

Conclusion

As part of the study, an autoencoder model was implemented and tested that solves the problem of automatic colorization of black and white images. The dataset served as the basis for the experiments *Landscape color and grayscale images*, containing over 14,000 landscape images. Using a relatively simple architecture with the Keras and TensorFlow libraries, we were able to create a reproducible model that delivers satisfactory results at a moderate computational cost.

Experiments have shown that the proposed model achieves an accuracy of about 82.5% on the validation set. Additional image quality metrics - PSNR (≈27.6 dB) and SSIM (≈0.83) - confirm the preservation of both numerical similarity and structural characteristics of the original images. Although the results are inferior to state-of-the-art architectures (U-Net, GAN), they demonstrate that even the basic configuration of the neural network is able to successfully solve the colorization problem and provide visually acceptable image quality.

The scientific novelty of the work lies in the practical implementation and testing of a simplified autoencoder, which, with limited resources, ensures reproducibility and stability of training, and also allows an objective assessment of the relationship between the complexity of the architecture and the quality of the result. The model is positioned as a resource-efficient basic solution, suitable for educational and research purposes, as well as for applied problems where a balance between quality and costs is required.

Prospects for further work are related to expanding the data set beyond landscape topics, implementing modern architectural approaches (U-Net, GAN, transformer models), and using additional quality metrics (LPIPS, FID) reflecting the perceptual features of image perception. A separate area of development is related to integrating the method into real applications — digital photo restoration systems, multimedia services, and designer tools.

Thus, the set tasks were completed, and the goal of the study was achieved: a working model of a neural network for image colorization was created and tested, demonstrating stable results and confirming the applicability of deep learning in the field of color information restoration.

REFERENCES

Anwar, S., Tahir, M., Li, C., Mian, A., Khan, F.S., Muzaffar, A.W. (2020). Image colorization: a survey and dataset [Image colorization: a survey and dataset]. — *arXiv preprint*, arXiv:2008.10774. — Pp. 1–14. Publisher: Cornell University Library (arXiv.org). [In Eng.]

Baldassarre, F., Polidori, L., Cavalieri, F., Russo, R. (2020). Auto colorization of historical images using deep convolutional neural networks [Auto colorization of historical images using deep CNNs]. — *Mathematics*. — Vol. 8(12). — Article 2258. — Pp. 1–18. Publisher: MDPI. [In Eng.]

Billaut, V., de Rochemonteix, M., Thibault, M. (2018). ColorUNet: a convolutional classification approach to colorization [ColorUNet: a CNN classification approach]. — *arXiv preprint*. — arXiv:1811.03120. — Pp. 1–10. Publisher: Cornell University Library (arXiv.org). [In Eng.]

Bogdanchikov, A., Ayazbayev, D., Varlamis, I. (2023). Classification of scientific documents in the Kazakh language using deep neural networks and a fusion of images and text [Classification of Kazakh scientific documents using DNNs]. — *Big Data and Cognitive Computing*. — Vol. 6(4). —Article 123. — Pp. 1–15. Publisher: MDPI. [In Eng.]

Cao, Y., Duan, X., Meng, X., Mok, P.Y., Li, P., Lee, T.Y. (2024). Computer-aided colorization state of the science: a survey [Computer-aided colorization: a survey]. — *arXiv preprint*, arXiv:2410.02288. — Pp. 1–30. Publisher: Cornell University Library (arXiv.org). [In Eng.]

Charpiat, G., LeCun, Y., Harchaoui, Z. (2008). Automatic image colourisation using SVM [Automatic image colourisation using SVM]. — In: *Proceedings of ECCV 2008*. — Pp. 1–6. Publisher: Springer. [In Eng.]

Geeks for Geeks. (2022). Colorization autoencoders using Keras [Colorization autoencoders using Keras]. — *GeeksforGeeks*. Publisher: GeeksforGeeks Pvt Ltd. Retrieved from https://www.geeksforgeeks.org. [In Eng.]

GitHub Contributors. (2021). Image colorization using autoencoders [Image colorization using autoencoders]. — *International Journal for Research in Applied Science & Engineering Technology*. — Vol. 9(9). — Pp. 1679–1683. Publisher: IJRASET. [In Eng.]

Goldashova, K., Sabdenov, Y., Uskenbayeva, G. (2022). Automatic colorization of Kazakh raster maps using CNNs [Avtomatty stylizatsiya Kazakh kartalar CNN arqyly {Автоматическая колоризация казахских растровых карт с использованием CNN}]. — *Journal of Eurasian Computing*. — Vol. 14(3). — Pp. 45–54. Publisher: Eurasian Computing Society. [In Kaz.]

Kaldarova, M., Akanova, A., Naizagarayeva, A., Kazanbayeva, A., Ospanova, N. (2022). Modelling a neural network for analysing the results of segmentation of satellite images [Model for satellite image segmentation]. — *Indonesian Journal of Electrical Engineering & Computer Science*. — Vol. 36(1). — Pp. 614–621. Publisher: Institute of Advanced Engineering and Science (IAES). [In Eng.]

Khassenova, B., Kulmagambetova, S. (2023). Colorization of historical photographs from Kazakhstan: a deep learning approach [Tarikhiy surattar kolorizatsiyasy {Колоризация исторических фотографий Казахстана}]. — *Central Asian Journal of Computer Science*. — Vol. 2(1). — Pp. 12–20. Publisher: Central Asian Academic Press. [In Kaz.]

Ma, X., Liu, W., Zhao, C., Tukhvatulina, L.R. (2024). Can large language model predict employee attrition? [Can LLM predict employee attrition?]. — *arXiv preprint*, arXiv:2411.01353. — Pp. 1–12. Publisher: Cornell University Library (arXiv.org). [In Eng.]

Mullins, C. (2022). DuckDB Wasm: fast analytical processing for the web [DuckDB Wasm: fast web analytics]. — *Proceedings of the VLDB Endowment*. — Vol. 15(11). — Pp. 2643–2646. Publisher: VLDB Endowment / ACM. [In Eng.]

Orazbayeva, G.M., Zhumagaliyeva, L.N. (2023). Intellektual'nyy analiz personala na osnove vektorov {Интеллектуальный анализ персонала на основе векторов} [Intelligent personnel analysis based on vectors]. — *Vestnik informatiki*. — Vol. 18(2). — Pp. 45–53. Publisher: ENU Press. [In Russ.]

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

Подписано в печать 15.09.2025.
Формат 60x881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59