

МРНТИ 28.23.37
УДК 004.032.26

<https://doi.org/10.55452/1998-6688-2025-22-2-67-75>

^{1*}Кожагулов Е.Т.,

PhD, и.о. доцента, ORCID ID: 0000-0001-5714-832X

*e-mail: kazgu.kz@gmail.com

¹Мақсұтова А.А.,

докторант, ORCID ID: 0000-0002-8601-8900

e-mail: sagalua95@gmail.com

¹Жексебай Д.М.,

PhD, ORCID ID: 0009-0008-1884-4662

e-mail: zhexebay92@gmail.com

¹Сқабылов А.А.,

PhD, ORCID ID: 0000-0002-5196-8252

e-mail: skabylov212@gmail.com

²Кожагулов Т.М.,

кан. пед. наук, профессор, ORCID ID: 0009-0001-3712-4459

e-mail: tokkozhagul@mail.ru

¹Казахский национальный университет им. аль-Фараби

²Казахская национальная академия искусств им. Жургенова

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СОВРЕМЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ ОБЪЕКТОВ ИСКУССТВ

Аннотация

В наше время проблема определения подлинности произведений искусства становится особенно актуальной. Ранее этот процесс выполнялся исключительно вручную, что требовало значительных затрат человеческих ресурсов. Однако с развитием технологий возникла потребность в создании интеллектуальной системы, способной проводить точный анализ и идентификацию произведений искусства. Уже существуют исследования, посвященные распознаванию объектов с различными топологиями на основе искусственного интеллекта, которые комбинируют методы глубокого обучения и машинного зрения [1–3]. Распознавание объектов искусства с использованием нейронных сетей – это область компьютерного зрения, направленная на идентификацию и классификацию произведений искусства, таких как картины, скульптуры и артефакты, с применением методов глубокого обучения. В крупнейших музеях и мировых хранилищах, таких как музей Лувр (Франция), Метрополитен-музей (США), Галерея Уффици (Италия), Британский музей (Великобритания), Рейксмузеум (Нидерланды), уже активно применяются современные технологии аутентификации произведений искусства, основанные на искусственном интеллекте и нейронных сетях. Эти системы позволяют значительно повысить точность определения подлинности экспонатов, автоматизировать процесс их анализа и минимизировать влияние человеческого фактора. Современные системы аутентификации, внедренные в крупнейшие музеи мира, позволяют не только более точно выявлять подлинность произведений искусства, но и защищать культурное наследие от фальсификаций и разрушения. Сверточные нейронные сети (CNN) – анализируют визуальные особенности картины, включая мазки, текстуру поверхности и цветовые сочетания. Глубокое обучение – обученные модели сравнивают исследуемое произведение с базой подлинных картин, выявляя потенциальные несоответствия. В данной статье представлено сравнительное исследование различных моделей нейронных сетей, направленное на повышение точности распознавания и обработки данных.

Ключевые слова: нейронные сети, распознавание, матрица запутанности, объекты искусства.

Введение

Развитие технологий искусственного интеллекта значительно повлияло на область распознавания изображений, включая анализ произведений искусства. Современные нейронные сети демонстрируют высокий уровень точности при классификации изображений, что делает их востребованными в музеях, галереях и исследовательских центрах. Однако выбор оптимальной архитектуры для данной задачи остается актуальной проблемой [4].

Существуют различные подходы к распознаванию объектов искусства с использованием нейронных сетей. В частности, сверточные нейронные сети (CNN), такие как ResNet, продемонстрировали высокую точность при классификации изображений благодаря глубокой архитектуре и остаточным связям [5, 6].

Сверточные нейронные сети (CNN) являются наиболее распространенным методом обработки изображений. Такие архитектуры, как ResNet, VGG, Inception и EfficientNet, успешно применяются для классификации картин по художникам или стилям [7, 8, 9].

Дополнительно рассматриваются модели, ориентированные на обнаружение объектов, такие как YOLO [10, 11, 12] которые были адаптированы для классификации изображений. YOLOv8 представляет собой улучшенную версию этой модели, объединяя высокую скорость работы с высокой точностью.

Несмотря на активное развитие этих методов, остается вопрос о том, какая архитектура лучше всего подходит для распознавания произведений искусства. Настоящее исследование направлено на сравнительный анализ различных нейросетей в данной области.

Материалы и методы

До появления нейросетей и методов компьютерного зрения аутентификация картин проводилась вручную экспертами и с помощью различных научных методов. Эти подходы требуют высокой квалификации специалистов и значительных временных затрат. Традиционные методы аутентификации произведений искусства остаются важными, но имеют ряд недостатков: субъективность, трудоемкость и высокую стоимость. Интеграция нейросетей и компьютерного зрения позволяет автоматизировать многие процессы, снизить ошибки и повысить точность анализа. Преимущества нейронных сетей: сверточные нейронные сети (CNN) анализируют особенности изображения на уровне пикселей, выявляя характерные черты произведения. Сверточные нейронные сети – это особый тип нейронных сетей, разработанный для обработки двумерных данных, таких как изображения. Их ключевая особенность – способность автоматически извлекать важные признаки из изображения, минимизируя необходимость ручной обработки данных. CNN широко применяются в компьютерном зрении для задач классификации, детекции, сегментации и других видов анализа изображений.

В данном исследовании проведен сравнительный анализ современных нейронных сетей, применяемых для распознавания объектов искусства. Основное внимание уделено эффективности различных архитектур при классификации художественных произведений по авторам.

Для эксперимента были отобраны две популярные нейросетевые архитектуры:

- ♦ ResNet50 – глубокая сверточная нейросеть, широко используемая для задач классификации изображений.
- ♦ YOLOv8 – современная модель, ориентированная на детекцию объектов, но адаптированная для задачи классификации.

Сравнительный анализ характеристик двух моделей показан в таблице 1.

Таблица 1 – Сравнительный анализ характеристик двух моделей

Характеристика	YOLOv8	ResNet-50
Тип модели	Обнаружение объектов (Object Detection)	Классификация изображений (Image Classification)
Архитектура	CNN + PAFPN + Head	Глубокая сверточная сеть с остаточными блоками
Глубина сети	Зависит от версии (n/s/m/l/x)	50 слоев
Назначение	Обнаружение, сегментация, отслеживание объектов	Классификация изображений
Размер входного изображения	Обычно 640x640 (можно изменить)	Обычно 224x224
Выход	Координаты объектов, классы	Один класс (или вероятность по классам)
Поддержка сегментации	Да (YOLOv8-seg)	Нет напрямую
Применение на практике	Видеонаблюдение, автономные авто, медицина и т.д.	Распознавание изображений, извлечение признаков
Скорость работы (FPS)	Очень высокая (в зависимости от модели и устройства)	Быстрая, но не в реальном времени (на CPU)
Точность (mAP для detection)	~50–56% mAP (в зависимости от версии)	Не применяется (mAP не используется)
Предобученные веса	Да (на COCO и других датасетах)	Да (на ImageNet)
Преимущества	Ультрабыстрое обнаружение объектов	Устойчивая классификация, хорошее извлечение фич
Библиотеки и фреймворки	PyTorch, Ultralytics	PyTorch, TensorFlow, Keras
Дата релиза	2023	2015

В качестве исходных данных использовался набор изображений, содержащий 50 классов, соответствующих различным художникам. Данные были собраны из открытых источников и включали не менее 300 картин на каждый класс. Впоследствии для более точного анализа количество классов было сокращено до шести (по критерию наличия более 300 изображений в категории). В результате окончательный датасет включал 1800 изображений. Разделение данных для обучения составляет на обучающую (80%) и на тестовую (20%) выборки.

После завершения обучения производилось тестирование моделей на контрольной выборке. Анализировались показатели точности классификации, а также степень ошибки (recall, precision и F1-score) для каждого класса. Дополнительно была построена матрица запутанности, позволяющая выявить наиболее сложные для классификации группы художников.

Таким образом, предложенная методология обеспечивает объективное сравнение различных нейросетевых архитектур в задаче распознавания объектов искусства. В следующем разделе представлены полученные результаты и их анализ.

Результаты и обсуждение

С целью определения, какая из предложенных вариаций нейронной сети лучше, был проведен сравнительный анализ двух моделей, YOLO v8 и ResNet50. В качестве обучаемого набора данных была использована база данных 50 художников, с общим количеством 18 000 картин.

Матрица запутанности (confusion matrix) – это таблица, которая позволяет оценить качество работы модели машинного обучения, в частности нейронной сети, выполняющей задачу

классификации. На рисунке 1 показан рисунок тепловой карты запутанности модели нейронной сети ResNet50. Для тестовых данных были использованы картины 6 художников с наибольшим количеством картин (рисунок 1).

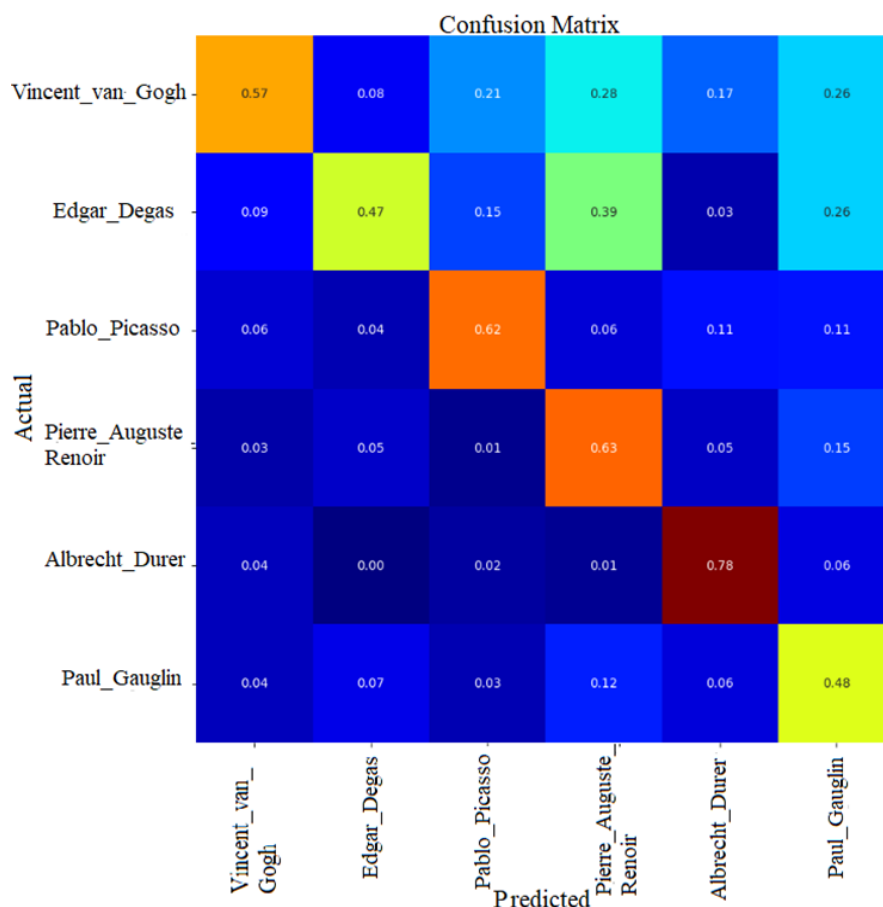


Рисунок 1 – Тепловая матрица запутанности модели ResNet50

Анализируя представленную матрицу запутанности, можно заключить, что нейросеть ResNet50 продемонстрировала недостаточно высокую точность классификации, что делает ее применение в данной задаче малопригодным. В среднем совпадение предсказанных и фактических классов составило около 60%, что явно недостаточно для точного распознавания картин.

Следующим этапом была использована нейронная сеть Yolo модель V8. YOLOv8 – это современная версия алгоритма YOLO (You Only Look Once), предназначенная для детекции и классификации объектов на изображениях. В отличие от ResNet50, которая больше ориентирована на распознавание и классификацию изображений в целом, YOLOv8 изначально создана для задач детекции, но ее можно адаптировать для классификации, особенно в задачах, связанных с картинами и искусством. На рисунке 2 представлена тепловая карта матрицы запутанности модели, обученной на наборе данных, включающем работы 50 художников. Обучение проводилось в течение 200 эпох.

Из рисунка 2 следует, что модель в большинстве случаев классифицирует правильно, но около 5,85% предсказаний ошибочны. Далее, для повышения точности распознавания количество категорий было сокращено до шести, включив лишь тех художников, чьи коллекции в наборе данных содержали более 300 картин. Этот отбор был сделан с целью более качественного обучения нейронной сети.

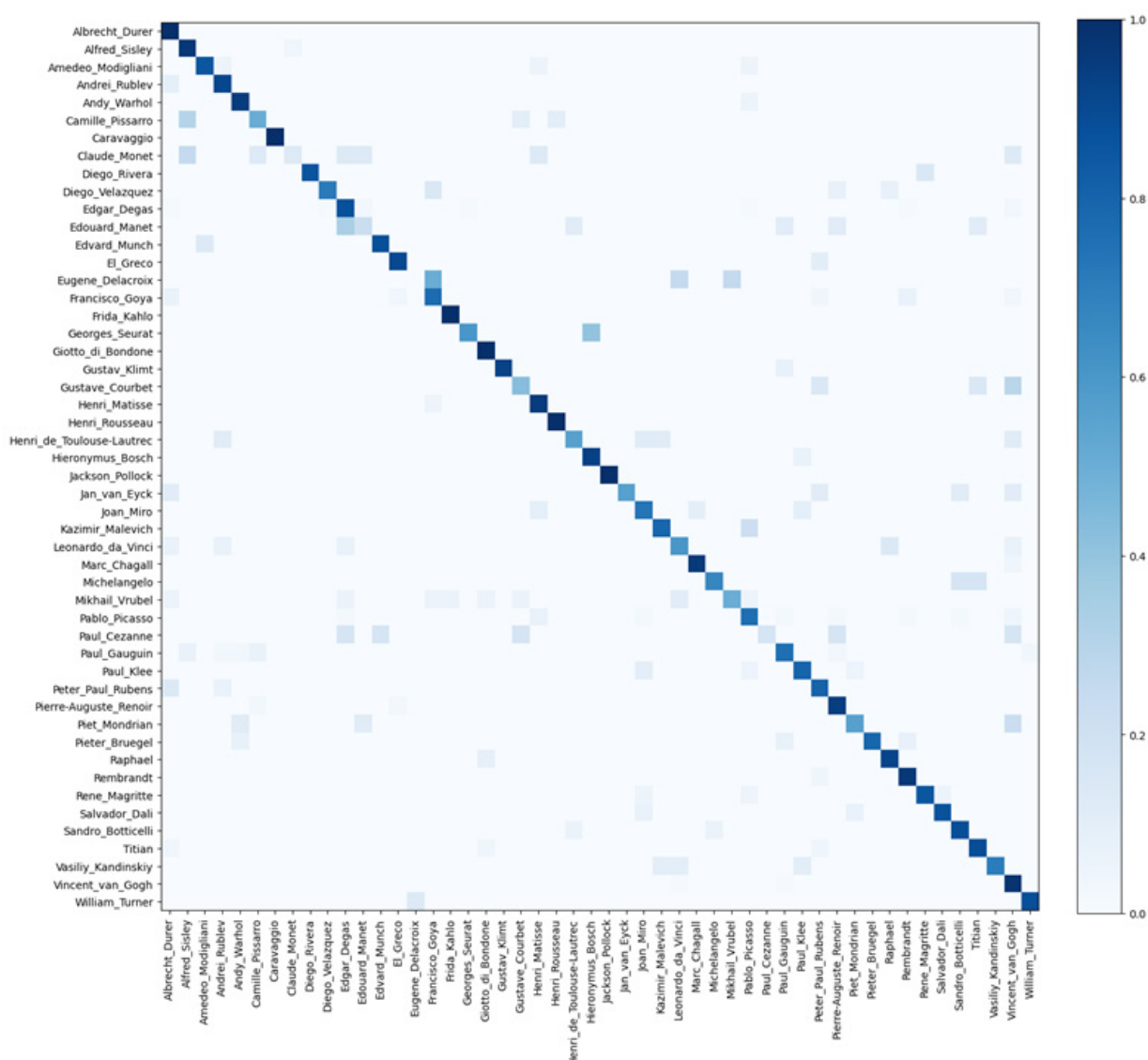


Рисунок 2 – Тепловая матрица запутанности модели Yolo V8

Таким образом, как представлено на рисунке 3, сокращение объема данных для более точного обучения модели привело к улучшению результатов. В общей сложности было использовано 1800 картин.

Полученные показатели обучения составили [0.86667, 0.86667, 0.93333, 0.93333, 0.96667, 0.96667]. Эти значения соответствуют диагональным элементам нормированной матрицы запутанности и отражают точность классификации для каждого из шести классов, представляющих работы различных художников.

Для последующего сравнения в данной модели количество эпох было увеличено до 500. Однако это не привело к значительным изменениям в результатах, при этом исходные данные обучения остались неизменными.

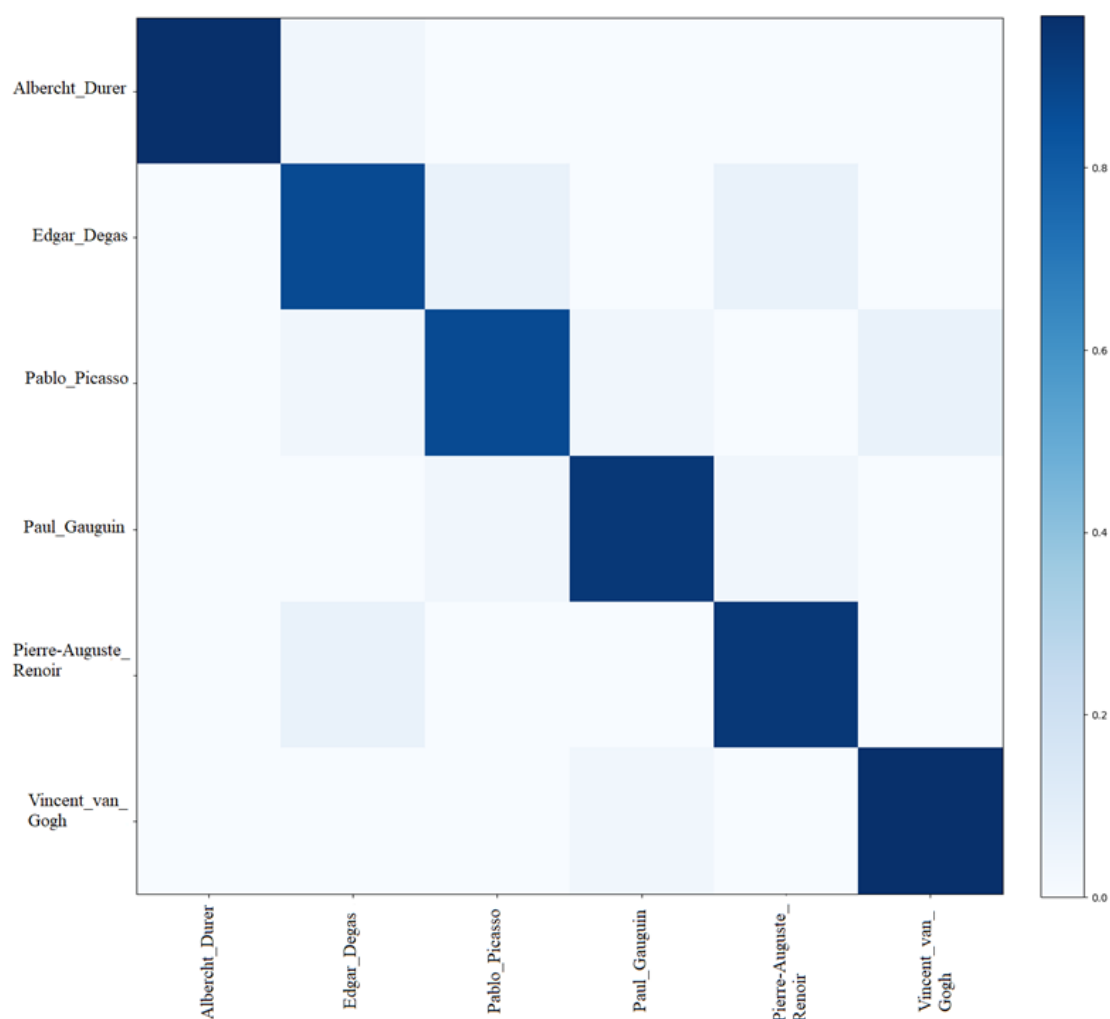


Рисунок 3 – Тепловая матрица запутанности модели YoloV8 с сокращенным набором данных

Заключение

В ходе проведенного исследования был выполнен сравнительный анализ современных нейронных сетей для распознавания произведений искусства. Были рассмотрены три архитектуры: ResNet50, Vision Transformer (ViT) и YOLOv8, каждая из которых показала свои особенности и преимущества в данной задаче.

Анализ результатов показал, что Vision Transformer продемонстрировал наибольшую точность при наличии большого объема данных, тогда как ResNet50 показал стабильные результаты с меньшими вычислительными затратами. YOLOv8, несмотря на свою изначальную направленность на детекцию объектов, также продемонстрировал высокий уровень классификации, особенно при оптимизации параметров обучения.

Дополнительные эксперименты, включающие увеличение количества эпох до 500, не привели к значительным улучшениям, что указывает на возможное перенасыщение модели. Использование аугментации данных и отбор наиболее представленных классов позволили достичь лучших результатов.

Таким образом, данное исследование подтверждает эффективность различных нейросетевых подходов к задаче распознавания произведений искусства и предоставляет рекомендации по выбору архитектуры в зависимости от доступных вычислительных ресурсов и объема дан-

ных. В дальнейшем возможно расширение исследования за счет включения более сложных ансамблевых методов и анализа стиливых особенностей произведений.

ЛИТЕРАТУРА

- 1 Кожугулов Е.Т., Жексебай Д.М., Сарманбетов С.А., МаксUTOва А.А., Бажаев Н.А. Классификатор изображений микросхем при помощи сверточной нейронной сети // News of the National Academy of Sciences of the Republic of Kazakhstan. Physico-Mathematical Series. – 2021. – Т. 6. – № 340. – С. 59–65. <https://doi.org/10.32014/2020.2518-1726.102/>
- 2 Kozhagulov Y.T., Zhexebay D.M., Sarmanbetov S.A., Sagatbayeva A.A., Zholdas D. Comparative analysis of object detection processing speed on the basis of neuroprocessors and neuroaccelerators // News of the National Academy of Sciences of the Republic of Kazakhstan. Series physico-mathematical. – 2020. – Vol. 4. – No. 332. – P. 61-67. <https://doi.org/10.32014/2020.2518-1726.66>
- 3 Kozhagulov Y.T., Ibraimov M.K., Zhexebay D.M., Sarmanbetov S.A. Face Detection of Integral Image by Viola-Jones method // Recent Contributions to Physics. – 2018. – No. 3 (66).
- 4 Simonyan K. and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. – 2014. <https://arxiv.org/abs/1409.1556>
- 5 Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. Microsoft Research. – 2016. <https://arxiv.org/abs/1512.03385>.
- 6 Jiazhi Liang. Image classification based on RESNET // Journal of Physics Conference Series. – 2020. – Vol.1634. – No. 1. – P. 012110. September 2020. <http://dx.doi.org/10.1088/1742-6596/1634/1/012110>
- 7 Szegedy Ch., Wei Liu, Yangqing Jia, Sermanet P., Reed S., Arbor A., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594>
- 8 Tan M. and Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, ICML 2019 (Long Beach, 9–15 June 2019), pp. 6105–6114. <http://proceedings.mlr.press/v97/tan19a.html>
- 9 Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Xiaohua Zhai, Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit Ja., Hounsby N. An image is worth 16x16 words: transformers for image recognition at scale. Published as a conference paper at ICLR 2021. <https://openreview.net/forum?id=YicbFdNTTy>
- 10 YOLOv8 Architecture: A Deep Dive into its Architecture. – 2024. YOLOv8.org. URL: <https://yolov8.org/yolov8-architecture/>
- 11 What is YOLO? The Ultimate Guide. Roboflow. – 2024. URL: <https://blog.roboflow.com/guideto-yolo-models>
- 12 Redmon Jo., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // Conference: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – June 2016. <http://dx.doi.org/10.1109/CVPR.2016.91>

REFERENCES

- 1 Kozhagulov E.T., Zheksebaj D.M., Sarmanbetov S.A., Maksutova A.A., Bazhaev N.A. Klassifikator izobrazhenij mikroschem pri pomoshhi svertochnoj nejronnoj seti. News of the National Academy of Sciences of the Republic of Kazakhstan. Physico-Mathematical Series, 6, 40, 59–65 (2021) <https://doi.org/10.32014/2020.2518-1726.102>.
- 2 Kozhagulov Y.T., Zhexebay D.M., Sarmanbetov S.A., Sagatbayeva A.A., Zholdas D. Comparative analysis of object detection processing speed on the basis of neuroprocessors and neuroaccelerators. News of the National Academy of Sciences of the Republic of Kazakhstan, Series physico-mathematical, 4 (332), 61–67 (2020). <https://doi.org/10.32014/2020.2518-1726.66>
- 3 Kozhagulov Y.T., Ibraimov M.K., Zhexebay D.M., Sarmanbetov S.A. Face Detection of Integral Image by Viola-Jones method. Recent Contributions to Physics, 3 (66) (2018).
- 4 Simonyan K. and Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition (2014). <https://arxiv.org/abs/1409.1556>
- 5 Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. Microsoft Research (2016). <https://arxiv.org/abs/1512.03385>

6 Jiazhi Liang. Image classification based on RESNET. Journal of Physics Conference Series, 1634(1), 012110 (September 2020). <http://dx.doi.org/10.1088/1742-6596/1634/1/012110>

7 Szegedy Ch., Wei Liu, Yangqing Jia, Sermanet P., Reed S., Arbor A., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A.. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594>

8 Tan M. and Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, ICML 2019 (Long Beach, 9–15 June 2019), pp. 6105–6114. <http://proceedings.mlr.press/v97/tan19a.html>

9 Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Xiaohua Zhai, Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit Ja., Houlsby N. An image is worth 16x16 words: transformers for image recognition at scale. Published as a conference paper at ICLR 2021 <https://openreview.net/forum?id=YicbFdNTTy>

10 YOLOv8 Architecture: A Deep Dive into its Architecture (2024). URL: <https://yolov8.org/yolov8-architecture/>

11 What is YOLO? The Ultimate Guide. Roboflow (2024). URL: <https://blog.roboflow.com/guideto-yolo-models>.

12 Redmon Jo., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. Conference: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016). <http://dx.doi.org/10.1109/CVPR.2016.91>

^{1*}Кожажулов Е.Т.,

PhD, доцент м.а., ORCID ID: 0000-0001-5714-832X

*e-mail: kazgu.kz@gmail.com

¹Мақсұтова А.А.,

докторант, ORCID ID: 0000-0002-8601-8900

e-mail: sagalua95@gmail.com

¹Жексебай Д.М.,

PhD, ORCID ID: 0009-0008-1884-4662

e-mail: zhexebay92@gmail.com

¹Скабылов А.А.,

PhD, ORCID ID: 0000-0002-5196-8252

e-mail: skabylov212@gmail.com

²Кожажулов Т.М.

пед.ғ.к., профессор

ORCID ID: 0009-0001-3712-4459

e-mail: tokkozbagul@mail.ru

¹эл-Фараби атындағы Қазақ ұлттық университет, Алматы қ., Қазақстан

²Темірбек Жүргенов атындағы қазақ ұлттық өнер академиясы, Алматы қ., Қазақстан

ӨНЕР ТУЫНДЫЛАРЫН ТАЛУҒА АРНАЛҒАН ЗАМАНАУИ НЕЙРОНДЫҚ ЖЕЛІЛЕРДІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ

Аңдатпа

Қазіргі таңда өнер туындыларының шынайылығын анықтау мәселесі ерекше өзекті. Бұрын бұл процесс тек қолмен жүзеге асырылатын, бұл айтарлықтай адам ресурстарын қажет ететін. Дегенмен, техниканың дамуымен өнер туындыларын дәл талдау мен сәйкестендіруге қабілетті интеллектуалды жүйені құру қажеттілігі туындады. Терең оқыту мен машиналық көру әдістерін біріктіретін жасанды интеллект негізінде әртүрлі топологиясы бар объектілерді тануға арналған зерттеулер қазірдің өзінде бар [1–3]. Нейрондық өнерді тану – терең оқыту әдістерін пайдалана отырып, кескіндеме, мүсіндер және артефактілер сияқты өнер туындыларын анықтауға және жіктеуге бағытталған компьютерлік көру саласы. Лувр мұражайы (Франция), Метрополитен мұражайы (АҚШ), Уффици галереясы (Италия), Британ мұражайы (Ұлыбритания) және Рейксмузеум (Нидерланды) сияқты ірі мұражайлар мен әлемдік репозиторийлер жасанды интеллект негізіндегі өнер туындыларының аутентификациясы үшін заманауи технологияларды белсенді түрде

пайдалануда. Бұл жүйелер экспонаттардың түпнұсқалығын анықтаудың дәлдігін айтарлықтай арттыруға, оларды талдау процесін автоматтандыруға және адам факторының әсерін барынша азайтуға мүмкіндік береді. Әлемнің ең ірі мұражайларында енгізілген заманауи аутентификация жүйелері өнер туындыларының түпнұсқалығын дәлірек анықтауға ғана емес, мәдени мұраны бұрмалау мен жойылудан қорғауға да мүмкіндік береді. Конволюциялық нейрондық желілер (CNN) кескіндеменің визуалды мүмкіндіктерін, соның ішінде шеткаларды, беттік текстураны және түс комбинацияларын талдайды. Терең оқытуда дайындалған модельдер зерттелетін жұмысты шынайы картиналар базасымен салыстырып, ықтимал сәйкессіздіктерді анықтайды. Бұл мақалада деректерді тану мен өңдеудің дәлдігін арттыруға бағытталған әртүрлі нейрондық желі модельдерінің салыстырмалы зерттеуі берілген.

Тірек сөздер: нейрондық желілер, тану, шатастыру матрицасы, өнер объектілері.

¹*Kozhagulov Y.,

PhD, Acting Associate Professor, ORCID ID: 0000-0001-5714-832X

*e-mail: kazgu.kz@gmail.com

¹Maksutova A.,

PhD student, ORCID ID: 0000-0002-8601-8900

e-mail: sagalua95@gmail.com

¹Zhexebay D.,

PhD, ORCID ID: 0009-0008-1884-4662

e-mail: zhexebay92@gmail.com

¹Skabylov A.,

PhD, ORCID ID: 0000-0002-5196-8252

e-mail: skabylov212@gmail.com

²Kozhagulov T.

Cand. Ped. Sci., Professor

ORCID ID: 0009-0001-3712-4459

e-mail: tokkozhaugul@mail.ru

¹Farabi Kazakh National University

²Kazakh National Academy of Arts named after T.K. Zhurgenov

COMPARATIVE ANALYSIS OF STATE-OF-THE-ART NEURAL NETWORKS FOR ART OBJECT RECOGNITION

Abstract

Currently, information technology is rapidly developing and one of its branches can be called machine translation. The use of machine translation in the process of understanding each other by people from different countries is increasing every year. At the moment, Google and Yandex machine translations are among the best machine translations. The quality of machine translation from Yandex and Google is improving every year. However, according to the results of the experiment, when translating from English or Russian into Kazakh and Turkic languages, the quality of the translation decreases. This was shown by the translation result obtained from these two machine translations in March 2024. After all, translation has also shown that it is directly related to the structure of language. Since 2000, scientists from the state of Kazakhstan have been actively studying translations into the Kazakh language. The goal of the work is to improve the quality of translation from English into Kazakh. For this purpose, a transforming model was created for the Kazakh and Turkic languages for learning translation in neural machine translation OpenNMT(). The created model studied and learned an English-Kazakh parallel corpus of 180,000 words. Later, the document with a structure of 20,000 different English sentences was translated into Kazakh. The result is measured using the Blue() metric. The translation result showed a high level. It is shown that in order to improve the results of the experiment carried out in the work during model training, it is necessary to increase the number of parallel corpora created from the English-Kazakh language pair.

Keywords: neural networks, recognition, confusion matrix, art objects.

Дата поступления статьи в редакцию: 19.02.2025