З.Г. ХАНАМИРЯН, Е.М. ТЕРТЕРЯН

МОДЕЛИРОВАНИЕ РОБОТА-МАНИПУЛЯТОРА В ПРОГРАММНОЙ СРЕДЕ LABVIEW

Роботы-манипуляторы стали неотъемлемой частью нашей повседневной жизни. С каждым днем растет их применение в промышленности, быту, медицине и других сферах. До производственного этапа роботов-манипуляторов необходимо разработать и изучить их симуляционную модель. Поэтому в среде LabVIEW Robotics было разработано программное приложение, которое позволяет изучать роботы-манипуляторы разных размеров и их рабочее пространство.

***Ключевые слова**:* робот, робототехника, кинематика, манипулятор, LabVIEW.

Z.G. KHANAMIRYAN, E.M. TERTERYAN

SIMULATION OF A ROBOT MANIPULATOR IN THE LABVIEW SOFTWARE ENVIRONMENT

Robot manipulators have become an integral part of our daily life. Their use in industry, everyday life, medicine, and other fields is growing every day. Before the production stage of robot manipulators, it is necessary to develop and study their simulation model. Therefore, a software application has been developed in the LabVIEW Robotics environment which allows to study robot manipulators of different sizes, and their workspace.

Keywords: robot, robotics, kinematics, manipulator, LabVIEW.

UDC 004.832

K.H. NIKOGHOSYAN, T.B. KHACHATRYAN, E.A. HARUTYUNYAN, D.M. GALSTYAN

EVALUATING OPEN-SOURCE IMAGE CAPTIONING MODELS WITH MULTIPLE METRICS ON THE IAPR TC-12 DATASET

In recent years, the development of image captioning AI models has been a focal point in the fields of computer vision and natural language processing (NLP). The paper presents a thorough comparative analysis of several state-of-the-art image captioning AI models, employing a diverse array of evaluation metrics, including CIDEr-D, BLEU-4, METEOR, ROUGE-L, SPICE, and Wu-Palmer similarity. The study is centered on the evaluation of image captioning models using the IAPR TC-12 dataset, a well-established benchmark for assessing visual content understanding. By leveraging multiple evaluation metrics, it was possible to gain a multifaceted understanding of the models' performance, encompassing both syntactic and semantic dimensions of generated captions. Comparative

analysis highlights that different metrics capture distinct facets of image captioning quality with each shedding light on specific aspects of model performance.

In summary, this paper offers a valuable resource for researchers in the fields of computer vision and natural language processing. This comprehensive assessment of image captioning models using multiple evaluation metrics and the IAPR TC-12 dataset provides a deeper understanding of the current capabilities and limitations of AI-driven approaches for generating descriptive image captions. This analysis paves the way for future advancements in this rapidly evolving domain.

Keywords: Artificial Intelligence, image captioning, natural language processing, evaluation metrics, computer vision.

Introduction. In the digital age, it is increasingly important to understand how text and images relate to each other. It plays an important role in computer vision, NLP, and many other applications. Computer vision and NLP are fundamental technologies that are changing and revolutionizing daily lives [1].

These techniques enable computers to understand and interact with the world by detecting objects, recognizing faces, and summarizing documents. A more comprehensive understanding of images, videos, and content will enable consumers to recognize, describe and provide automated content recommendations, thereby increasing efficiency and enabling new applications across a range of industries. It's important to understand how text and images work together when discussing social media, where most posts contain both text and visuals. As well as enabling us to search for relevant images based on text queries, it also powers search engines.

Many models have already been created that solve the problem of obtaining descriptions from images. In those models, there are some challenges, such as understanding context, handling unique items, bias issues, scalability, multimodal integration, privacy etc. To solve these problems, image captions in various applications need to be improved.

Related Works. A study [2] proposed a method that generated image descriptions closer to human speech. A new method, Scene Graph Auto-Encoder (SGAE), has been developed to teach the computer this style without the need for many examples. Extensive testing has shown that this approach has performed well, outperforming current methods on MS-COCO dataset images.

In [3], the application of generating descriptions from images in the field of fashion is considered. The researchers proposed a new way of learning captions and created a preliminary dataset of fashion captions called FACAD. In order to accurately and expressively describe fashion items, two new measurements have been defined: ALS and SLS. The model was then trained using these metrics along with MLE, feature embedding, and reinforcement learning. The paper also presents general measurements for the estimation of captions for common images. The

authors also noted that further research is needed to improve the evaluation criteria.

[4] proposes an extension of the LSTM model to generate captions for images. Unlike other methods, in this one, the authors proposed to include additional semantic information with each unit of the LSTM block. As a result, it was shown that the model can maintain a more accurate description of the image content and avoid drifting into unrelated regular expressions. The proposed method in this study achieves state-of-the-art performance on various benchmark datasets.

Materials and methods. Instead of teaching the model all about images and language, which can be time-consuming and computationally expensive, pre-trained models were used. This approach significantly speeds up the process of creating captions for images. Below is the list of those models:

- BLIP (Bidirectional Language Model Pretraining) [5];
- Image Captioning with PyTorch [6];
- Bottom-Up Attention for Image Understanding [7];
- Prismer [8];
- Adaptive Attention for Image Captioning [9];
- Neuraltalk2 [10];
- Image-Captioning-Transformer [11].

Here's how it works. An input image is fed to a pre-trained model, which quickly analyzes the image and provides a descriptive caption based on what it has learned from many other images. All the above models were trained on the popular COCO dataset, as well as on several relatively small datasets.

Dataset. For experiments, the IAPR TC-12 dataset was used for evaluating the already pre-trained model outputs. The dataset includes 20,000 natural images (Fig. 1) accompanied by subtitles in three languages: English, German, and Spanish. It includes a wide variety of topics, such as sports, people, animals, cities, and landscapes. This dataset has been utilized by researchers for many different tasks, such as semantic indexing, cross-modal search, and image captioning. It has also been expanded to include segmentation and annotation data, which makes it easier to assess the automatic image annotation methods [12].

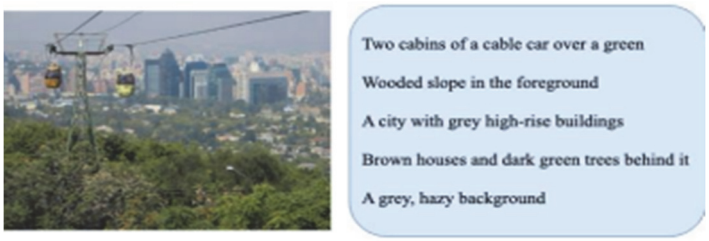

Fig. 1. IAPR TC-12 dataset image and label example

Evaluation metrics. The image-to-description task initially faced challenges due to the gap between visual and textual data, but semantic analysis now plays a crucial role in solving this problem. Understanding the meaning of the text has challenges. Things like language details, context, and unclear meanings make it hard. Also, dealing with various text sources, like different languages or topics, makes it even more complex to get the right and clear meanings.

The evaluation of image-capturing models was conducted using six metrics. In this section, they are all described.

- CIDEr-D: measures the similarity between a candidate caption and a set of reference captions based on the n-gram overlap:

$$CIDEr - D(c,S) = \frac{10}{M}\sum_{m-1}^{M} \quad \sum_{n=1}^{N} \quad w_n sim_n(c,s_m), \tag{1}$$

where c is the candidate caption, S - the set of reference captions, M - the number of reference captions, N - the maximum n-gram order (usually 4), w_n - the weight for each n-gram order, and $sim_n(c,s_m)$ - the n-gram similarity between c and s_m.

- BLEU measures the precision of a candidate caption with respect to one or more reference captions based on the n-gram overlap.

$$BLEU(c,S) = BP * exp(\sum_{n=1}^{N} \quad w_n logp_n), \tag{2}$$

where c is the candidate caption, S - the set of reference captions, BP - the brevity penalty, N - the maximum n-gram order (usually 4), w_n - the weight for each n - gram order, and p_n - the precision for each n-gram order.

- METEOR measures the harmonic mean of unigram precision and recall between a candidate caption and a reference caption based on exact, stem, synonym, and paraphrase matches. The formula for METEOR is:

$$METEOR(c,r) = (1-\alpha)P + \alpha R, \tag{3}$$

where c is the candidate caption, r - the reference caption, α - a parameter that controls the balance between precision and recall (usually 0.85), P - the unigram precision, and R the unigram recall.

- ROUGE-L measures the longest common subsequence (LCS) between a candidate caption and a reference caption. The formula for ROUGE-L is:

$$ROUGE - L(c,r) = \frac{(1+\beta^2)RP}{R+\beta^2 P} \tag{4}$$

where c is the candidate caption, r - the reference caption, β^2 - a parameter that controls the balance between precision and recall, R - the LCS-based recall, and P - the LCS-based precision.

• SPICE computes the F-measure of scene graph tuple precision and recall reflecting the quality and quantity of semantic matches. The formula for SPICE is:

$$SPICE(c,r) = \frac{(1+\lambda^2)RP}{R+\lambda^2 P}, \quad (5)$$

where c is the candidate caption, r - the reference caption, λ^2 - a parameter that controls the balance between precision and recall (usually 1), R - the scene graph tuple recall, and P - the scene graph tuple precision.

● NLTK (Natural Language Toolkit) and WordNet enable the comparison of sentence meanings based on their semantic similarity. WordNet is a lexical database that provides sets of synonyms (synsets) for words, which can be used to construct a similarity measure between sentences. The following steps outline how to use NLTK and WordNet for this purpose:

o the first step is to tokenize and preprocess the sentences, which involves transforming the sentences into word lists, removing punctuation, converting to lowercase, and applying any other required preprocessing steps;

o the next step is to find the WordNet synsets for the words in the sentences using NLTK's WordNet interface;

o the final step is to calculate a semantic similarity measure between the sentences using the synsets. A common measure is the Wu Palmer Similarity, which computes the similarity between each pair of synsets from the two sentences and records the highest similarity score. This example illustrates a basic approach to sentence similarity using WordNet and NLTK. However, the results may not always be accurate because of the limitations of WordNet's coverage and sense disambiguation compared to more advanced methods such as word embeddings or transformer-based models.

The mathematical formula for Wu-Palmer Similarity is:

$$Wu - Palmer\ Similarity = \frac{2\times depth\ of\ LCS}{depth\ of\ synset\ 1 + depth\ of\ synset\ 2}, \quad (6)$$

where LCS stands for Least common subsumer, which is the lowest common ancestor of two synsets in the WordNet hierarchy, and depth is the number of edges from the root node to the synset.

Results and Discussion. These measures were used to assess and reflect the performance of several picture captioning algorithms in relation to one another. Each statistic evaluates a different element of caption quality, adding to a thorough

assessment of the models on the IAPR TC-12 dataset. The table below shows the metric measures:

Table

Evaluation of image captioning algorithms on IAPR TC-12 dataset

Model	CIDEr-D	BLEU-4	METEOR	ROUGE-L	SPICE	Wu-Palmer
PyTorch Image Captioning	1.204	0.357	0.280	0.569	0.216	0.538
Bottom-Up Top Down	1.201	0.363	0.277	0.560	0.205	0.519
Neuraltalk2	0.855	0.277	0.233	0.516	0.203	0.431
Image-Captioning Transformer	1.143	0.346	0.269	0.554	0.206	0.525
BLIP	1.179	0.369	0.245	0.531	0.215	0.496
Prismer	1.206	0.378	0.221	0.525	0.22	0.532
Adaptive Attention for Image Captioning	1.085	0.357	0.275	0.564	0.2	0.536

Conclusion. In summary, efficiency and quality are critical for picture captioning tasks. To investigate the process, researchers have turned to pre-trained models such as BLIP. These models, which have been extensively trained on datasets such as COCO, provide a shortcut to correct picture captioning. The IAPR TC-12 dataset, which has varied pictures and multilingual descriptions, was utilized to evaluate their performance**.** In image captioning experiments, multiple models were evaluated using various evaluation metrics. The PyTorch Image Captioning model excelled in CIDEr-D, METEOR, ROUGE-L, and SPICE metrics, indicating its overall strong performance in generating image captions. Meanwhile, the Bottom Up Top-Down model stood out in terms of BLEU-4, showcasing its precision in caption generation. The best model among the seven, according to the Wu-Palmer measure also is PyTorch Image Captioning, with a score of 0.538. This indicates that, unlike the others, this model may create captions that are semantically more comparable to the reference captions.

The evaluation conducted in this study serves as a valuable initial step in selecting a baseline model for image captioning tasks, while also contributing to the ongoing development of research in this domain.

REFERENCES

1. **El-Komy A., Shahin O.R., Abd El-Aziz R.M., & Taloba A.I.** Integration of computer vision and natural language processing in multimedia robotics application //Inf. Sci.- 2022.- 7(6).

2. **Yang X., Tang K., Zhang H., and Cai J.** Auto-encoding scene graphs for image captioning //IEEE/CVF Conf. Computer Vision and Pattern Recognition.- 2019.- P. 10685−10694.
3. Fashion captioning: Towards generating accurate descriptions with semantic rewards / **X. Yang, H. Zhang, D. Jin, et al** //Proc. Computer Vision-ECCV.- Springer, 2020.- P. 1−17.
4. **Jia X., Gavves E., Fernando B., and Tuytelaars T.** Guiding the long short term memory model for image caption generation //Proc. IEEE/CVF Int. Conf. Computer Vision.- 2015.- P. 2407−2415.
5. **Li J., Li D., Xiong C., & Hoi S.** Blip: Bootstrapping language-image pre training for unified vision-language understanding and generation //International Conference on Machine Learning.- 2022.- P. 12888-12900.
6. Empowering image captioning models with ownership protection / **J.H. Lim, C. S. Chan, K.W. Ng, et al** //Pattern Recognition.-2022.
7. **Pan Y., Li Y., Yao T., & Mei T.** Bottom-up and Top-down Object Inference Networks for Image Captioning //ACM Transactions on Multimedia Computing, Communications and Applications. – 2023.- P. 1-18.
8. **Prismer:** A vision-language model with an ensemble of experts / **S. Liu, et al.** - 2023.
9. Knowing when to look: Adaptive attention via a visual sentinel for image captioning /**J. Lu, et al** //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. - P. 375-383.
10. **Neuraltalk2** https://github.com/karpathy/neuraltalk2, Accessed: 29/09/23.
11. Discriminability objective for training descriptive captions /**R. Luo, et al** //Proceedings of the IEEE conference on computer vision and pattern recognition.- 2018.- P. 6964-6974.
12. The segmented and annotated IAPR TC-12 benchmark /**H.J. Escalante, et al** //Computer Vision and Image Understanding.- 2010.- P. 419-428.

Կ.Հ. ՆԻԿՈՂՈՍՅԱՆ, Տ.Բ. ԽԱՉԱՏՐՅԱՆ, Է.Ա. ՀԱՐՈՒԹՅՈՒՆՅԱՆ, Դ.Մ. ԳԱԼՍՏՅԱՆ

ԲԱՑ ՀԱՍԱՆԵԼԻՈՒԹՅԱՄԲ ՊԱՏԿԵՐԻ ՍՈՒԲՏԻՏՐԱՎՈՐՄԱՆ ՄՈԴԵԼՆԵՐԻ ԳՆԱՀԱՏՈՒՄ IAPR TC-12 ՏՎՅԱԼՆԵՐԻ ՀԱՎԱՔԱԾՈՒ ՎՐԱ

Վերջին տարիներին պատկերների սուբտիտրավորման արհեստական բանականության (ԱԲ) մոդելների զարգացումը ուշադրության կենտրոնում է եղել համակարգչային տեսողության և բնական լեզվի մշակման ոլորտներում: Աշխատանքում ներկայացվել է մի քանի ժամանակակից պատկերների սուբտիտրավորման ԱԲ մոդելների մանրակրկիտ համեմատական վերլուծություն՝ օգտագործելով գնահատման չափանիշների բազմաթիվ մեթոդներ, ինչպիսիք են՝ CIDEr-D, BLEU-4, METEOR, ROUGE-L, SPICE և Wu Palmer: Գնահատվել են պատկերների սուբտիտրավորման մոդելները՝ օգտագործելով IAPR TC-12 տվյալների հավաքածուն, որը լավ հաստատված չափանիշ է տեսողական բովանդակու-

թյան ըմբռնումը գնահատելու համար: Բազմաթիվ գնահատման չափումներ օգտագործելով՝ հնարավոր եղավ ձեռք բերել մոդելների կատարողականի բազմակողմանի պատկերացում՝ ներառելով ստեղծված ենթագրերի ինչպես շարահյուսական, այնպես էլ իմաստային չափումները:

Համեմատական վերլուծությամբ ընդգծվում է, որ տարբեր չափումներ ամրագրում են պատկերի սուբտիտրերի որակի տարբեր կողմերը, որոնցից յուրաքանչյուրը լուսաբանում է մոդելի կատարողականի որոշակի հայեցակետեր:

Աշխատանքն արժեքավոր ռեսուրս է առաջարկում հետազոտողների համար համակարգչային տեսողության և բնական լեզվի մշակման ոլորտներում: Պատկերների սուբտիտրավորման մոդելների այս համապարփակ գնահատումը, կատարելով բազմաթիվ չափումներ IAPR TC-12 տվյալների հավաքածուի վրա, ավելի խոր պատկերացում է տալիս նկարների սուբտիտրեր ստեղծելու համար ԱԲ-ի վրա հիմնված մոտեցումների ներկայիս հնարավորությունների և սահմանափակումների մասին: Վերլուծությունը արագ զարգացող այս ոլորտում ապագա առաջընթացի համար հիմք է ստեղծում:

Առանցքային բառեր. արհեստական բանականություն, պատկերների սուբտիտրեր, բնական լեզվի մշակում, գնահատման չափումներ, համակարգչային տեսողություն:

К.Г. НИКОГОСЯН, Т.Б. ХАЧАТРЯН, Э.А. АРУТЮНЯН, Д.М. ГАЛСТЯН

ОЦЕНКА МОДЕЛЕЙ СУБТИТРОВ К ИЗОБРАЖЕНИЯМ С ОТКРЫТЫМ ИСТОЧНИКОМ С НЕСКОЛЬКИМИ ПОКАЗАТЕЛЯМИ НА НАБОРЕ ДАННЫХ IAPR TC-12

В последние годы разработка моделей искусственного интеллекта для субтитров к изображениям была в центре внимания в области компьютерного зрения и обработки естественного языка. В статье представлен тщательный сравнительный анализ нескольких современных моделей искусственного интеллекта для субтитров с использованием разнообразного набора показателей оценки, включая CIDEr-D, BLEU-4, METEOR, ROUGE-L, SPICE и Wu-Palmer. Исследование сосредоточено на оценке моделей субтитров к изображениям с использованием набора данных IAPR TC-12, хорошо зарекомендовавшего себя эталона для оценки понимания визуального контента. Используя несколько показателей оценки, удалось получить многостороннее понимание производительности моделей, охватывающее как синтаксические, так и семантические аспекты генерируемых субтитров. Сравнительный анализ показывает, что разные показатели отражают разные аспекты качества субтитров к изображениям, причем каждый из них проливает свет на конкретные аспекты эффективности модели.

Статья предлагает ценный ресурс для исследователей в области компьютерного зрения и обработки естественного языка. Комплексная оценка моделей субтитров к изображениям с использованием нескольких показателей оценки и набора данных IAPR TC-12 обеспечивает более глубокое понимание текущих возможностей и

ограничений подходов на основе искусственного интеллекта для создания описательных субтитров к изображениям. Этот анализ прокладывает путь для будущих достижений в этой быстро развивающейся области.

Ключевые слова: искусственный интеллект, субтитры к изображениям, обработка естественного языка, метрики оценки, компьютерное зрение.

ՀՏԴ 621.01:631.3

Տ.Ա. ԲԱՐՍԵՂՅԱՆ, Ք.Գ. ԱՎԵՏԻՍՅԱՆ

ԳՅՈՒՂԱՏՆՏԵՍԱԿԱՆ ԽՈՐՀՐԴԱՏՎԱԿԱՆ ՀԱՄԱԿԱՐԳԻ ՄՇԱԿՈՒՄԸ ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՄԱՆ ԿԻՐԱՌՄԱՄԲ

(Գյումրի)

Մշակվել է խորհրդատվական համակարգ մեքենայական ուսուցման ալգորիթմների օգտագործմամբ: Այդ ալգորիթմները թույլ կտան վերլուծել տարբեր սենսորներից ստացվող տվյալները, մշակել դրանք և կատարել ճշգրիտ խորհրդատվություն գյուղատնտեսին՝ նշելով ուսումնասիրվող հողատիպում աճեցման համար նպատակահարմար և ոչ նպատակահարմար բուսատեսակների անվանումները: Ծրագրի իրականացման համար ուսումնասիրվել են մեքենայական ուսուցման մի շարք ալգորիթմներ, Python ծրագրավորման լեզվի տարբեր գրադարաններ, PostgreSQL տվյալների բազաների կառավարման համակարգը, ինչպես նաև կատարվել է նախնական տվյալների հավաքագրում և մշակում՝ օգտագործելով տարբեր աղբյուրներ (օրինակ, ՀՀ ագրոկլիմայական գոտիների հիմնական հողատիպերի և, դրանցում աճող բուսատեսակների տեղաբաշխման աղյուսակը):

Առանցքային բառեր. մեքենայական ուսուցում, արհեստական բանականություն, ալգորիթմ, խորհրդատվական համակարգ, բովանդակության վրա հիմնված ֆիլտրում, խմբավորում, տվյալների բազա, հող, բուսատեսակներ, գյուղատնտեսություն:

Ներածություն: Հաշվի առնելով, Երկիր մոլորակում ծառացած գլոբալ խնդիրները, ինչպիսիք են՝ կլիմայի փոփոխությունը, գլոբալ տաքացման պրոցեսները, պարենի անվտանգությունը և բնական ռեսուրսների աղտոտումն ու հողերի վատթարացումը, կենսական անհրաժեշտություն է առաջանում բնական ռեսուրսների մոնիտորինգի և տվյալների հավաքագրման վերաբերյալ: Մոնիտորինգային տվյալների վերլուծությունը նպատակաուղղված է բնական ռեսուրսների կայուն և արդյունավետ օգտագործմանը, տեղային խնդիրների բացահայտմանը և մեղմմանը: Հայաստանի Հանրապետությունում, որին որպես սակավահող երկիր բնորոշ է ուղղաձիգ գոտիականությունը, կարևորվում է հողային ռեսուրսների ֆիզիկական, քիմիական և կենսաբանական հատկությունների վերաբերյալ ճշգրիտ տվյալների հավաքագրումը: Վերջինս հիմք է հանդիսանում գյուղատնտեսության սեկտորում ճիշտ որոշումների կայացման, ֆերմերային տնտեսությունների կողմից