DAFTAR PUSTAKA

Amal, I.I., Widyantoro, D.H., Umam, A., 2020. MobileNet-based Neural Image Caption Model in Title Generation for Product's Images, in: 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA). Presented at the 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA), pp. 1–6. https://doi.org/10.1109/ICAICTA49861.2020.9428886

Anderson, P., Fernando, B., Johnson, M., Gould, S., 2016. SPICE: Semantic Propositional Image Caption Evaluation [WWW Document]. arXiv.org. URL https://arxiv.org/abs/1607.08822v1 (accessed 9.6.23).

Banerjee, S., Lavie, A., 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments, in: Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. Association for Computational Linguistics, Ann Arbor, Michigan, pp. 65–72.

Freitag, M., Al-Onaizan, Y., 2017. Beam Search Strategies for Neural Machine Translation, in: Proceedings of the First Workshop on Neural Machine Translation. pp. 56–60. https://doi.org/10.18653/v1/W17-3207

Gupta, S.C., Singh, N.R., Sharma, T., Tyagi, A., Majumdar, R., 2021. Generating Image Captions using Deep Learning and Natural Language Processing, in: 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). Presented at the 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1–4. https://doi.org/10.1109/ICRITO51393.2021.9596486

Hartatik, Al Fatta, H., Fajar, U., 2019. Captioning Image Using Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM), in: 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). Presented at the 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), pp. 263–268. https://doi.org/10.1109/ISRITI48646.2019.9034562

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition.

Hendrycks, D., Gimpel, K., 2016. Gaussian Error Linear Units (GELUs) [WWW Document]. arXiv.org. URL https://arxiv.org/abs/1606.08415v5 (accessed 9.15.23).

Hoxha, G., Melgani, F., Slaghenauffi, J., 2020. A New CNN-RNN Framework For Remote Sensing Image Captioning, in: 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS). Presented at the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), pp. 1–4. https://doi.org/10.1109/M2GARSS47143.2020.9105191

Jaknamon, T., Marukatat, S., 2022. ThaiTC:Thai Transformer-based Image Captioning, in: 2022 17th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). Presented at the 2022 17th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), pp. 1–4. https://doi.org/10.1109/iSAI-NLP56921.2022.9960246

Kolen, J.F., Kremer, S.C., 2001. Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies, in: A Field Guide to Dynamical Recurrent Networks. Presented at the A Field Guide to Dynamical Recurrent Networks, IEEE, pp. 237–243. https://doi.org/10.1109/9780470544037.ch14

Lin, C.-Y., 2004. ROUGE: A Package for Automatic Evaluation of Summaries, in: Text Summarization Branches Out. Association for Computational Linguistics, Barcelona, Spain, pp. 74–81.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P., 2015. Microsoft COCO: Common Objects in Context. https://doi.org/10.48550/arXiv.1405.0312

Mahadi, M.R.S., Arifianto, A., Ramadhani, K.N., 2020. Adaptive Attention Generation for Indonesian Image Captioning, in: 2020 8th International Conference on Information and Communication Technology (ICoICT). Presented at the 2020 8th International Conference on Information and Communication Technology (ICoICT), pp. 1–6. https://doi.org/10.1109/ICoICT49345.2020.9166244

Meister, C., Cotterell, R., Vieira, T., 2020. If beam search is the answer, what was the question?, in: Webber, B., Cohn, T., He, Y., Liu, Y. (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Presented at the EMNLP 2020, Association for Computational Linguistics, Online, pp. 2173–2185. https://doi.org/10.18653/v1/2020.emnlp-main.170

Mulyanto, E., Setiawan, E.I., Yuniarno, E.M., Purnomo, M.H., 2019. Automatic Indonesian Image Caption Generation using CNN-LSTM Model and FEEH-ID Dataset, in: 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA). Presented at the 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp. 1–5. https://doi.org/10.1109/CIVEMSA45640.2019.9071632

Mulyawan, R., Sunyoto, A., Muhammad, A.H., 2022. Automatic Indonesian Image Captioning using CNN and Transformer-Based Model Approach, in: 2022 5th International Conference on Information and Communications Technology (ICOIACT). Presented at the 2022 5th International Conference on Information and Communications Technology (ICOIACT), pp. 355–360. https://doi.org/10.1109/ICOIACT55506.2022.9971855

Mulyawan, R., Sunyoto, A., Muhammad, A.H.M., 2023. Pre-Trained CNN Architecture Analysis for Transformer-Based Indonesian Image Caption Generation Model. JOIV Int. J. Inform. Vis. 7, 487–493. https://doi.org/10.30630/joiv.7.2.1387

Nugraha, A.A., Arifianto, A., Suyanto, 2019. Generating Image Description on Indonesian Language using Convolutional Neural Network and Gated Recurrent Unit, in: 2019 7th International Conference on Information and Communication Technology (ICoICT). Presented at the 2019 7th International Conference on Information and Communication Technology (ICoICT), pp. 1–6. https://doi.org/10.1109/ICoICT.2019.8835370

O'Shea, K., Nash, R., 2015. An Introduction to Convolutional Neural Networks.

Papineni, K., Roukos, S., Ward, T., Zhu, W.-J., 2002. Bleu: a Method for Automatic Evaluation of Machine Translation, in: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Presented at the ACL 2002, Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, pp. 311–318. https://doi.org/10.3115/1073083.1073135

Periasamy, J.K., K, M.I., T, K.P., 2022. Generating image description aiding blind people interpretation, in: 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT). Presented at the 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), pp. 1–4.

https://doi.org/10.1109/IC3IOT53935.2022.9767967

Plummer, B.A., Wang, L., Cervantes, C.M., Caicedo, J.C., Hockenmaier, J., Lazebnik, S., 2016. Flickr30k Entities: Collecting Region-to-Phrase Correspondences for Richer Image-to-Sentence Models. https://doi.org/10.48550/arXiv.1505.04870

Sharma, H., Agrahari, M., Singh, S.K., Firoj, M., Mishra, R.K., 2020. Image Captioning: A Comprehensive Survey, in: 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and Its Control (PARC). Presented at the 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), pp. 325–328. https://doi.org/10.1109/PARC49193.2020.236619

Sudhakar, J., Iyer, V.V., Sharmila, S.T., 2022. Image Caption Generation using Deep Neural Networks, in: 2022 International Conference for Advancement in Technology (ICONAT). Presented at the 2022 International Conference for Advancement in Technology (ICONAT), pp. 1–3. https://doi.org/10.1109/ICONAT53423.2022.9726074

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the Inception Architecture for Computer Vision.

Tanti, M., Gatt, A., Camilleri, K.P., 2018. Where to put the Image in an Image Caption Generator. Nat. Lang. Eng. 24, 467–489. https://doi.org/10.1017/S1351324918000098

Tsang, S.-H., 2019. Review: Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015. Medium. URL https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c (accessed 9.11.23).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need.

Vedantam, R., Zitnick, C.L., Parikh, D., 2014. CIDEr: Consensus-based Image Description Evaluation [WWW Document]. arXiv.org. URL https://arxiv.org/abs/1411.5726v2 (accessed 9.6.23).

Vinyals, O., Toshev, A., Bengio, S., Erhan, D., 2015. Show and Tell: A Neural Image Caption Generator.

Wang, G., Lu, Y., Cui, L., Lv, T., Florencio, D., Zhang, C., 2022. A Simple yet Effective Learnable Positional Encoding Method for Improving Document Transformer Model, in: Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022. Presented at the Findings 2022, Association for Computational Linguistics, Online only, pp. 453–463.

Xin, R., Zhang, J., Shao, Y., 2020. Complex network classification with convolutional neural network. Tsinghua Sci. Technol. 25, 447–457. https://doi.org/10.26599/TST.2019.9010055

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R., Bengio, Y., 2016. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention.

Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629. https://doi.org/10.1007/s13244-018-0639-9