

DAFTAR PUSTAKA

- Ahmad, R.A., Azhar, M. & Sattar, H. 2022. An Image captioning algorithm based on the Hybrid Deep Learning Technique (CNN+GRU). *2022 International Conference on Frontiers of Information Technology (FIT)*. 2022 International Conference on Frontiers of Information Technology (FIT). hlm.124–129. Tersedia di <https://ieeexplore.ieee.org/document/10043039> [Accessed 5 Juni 2024].
- Alam, I. & Nabiilah, G. 2024. Indonesian multilabel classification using IndoBERT embedding and MBERT classification. *International Journal of Electrical and Computer Engineering (IJECE)*, 14: 1071–1078.
- Alammar, J. 2024. *The Illustrated Transformer*. Tersedia di <https://jalammar.github.io/illustrated-transformer/> [Accessed 5 Juni 2024].
- Anonim, 2023. *BERT Embeddings*. Tersedia di <https://tinkerd.net/blog/machine-learning/bert-embeddings/> [Accessed 5 Juni 2024].
- Biswas, S. & Sharma, S. 2021. Image Descriptor Generator Using Encoder-Decoder Model. *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. 2021 6th International Conference on Communication and Electronics Systems (ICCES). hlm.1144–1147. Tersedia di <https://ieeexplore.ieee.org/document/9489250> [Accessed 5 Juni 2024].
- Charu, S., Mishra, S.P. & Gandhi, T. 2020. Vision to Language: Captioning Images using Deep Learning. *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*. 2020 International Conference on Artificial Intelligence and Signal Processing (AISP). hlm.1–8. Tersedia di <https://ieeexplore.ieee.org/document/9073009> [Accessed 5 Juni 2024].
- Cheng, Y.-H., Chang, P.-C. & Kuo, C.-N. 2020. Convolutional Neural Networks Approach for Music Genre Classification. *2020 International Symposium on Computer, Consumer and Control (IS3C)*. 2020 International Symposium on Computer, Consumer and Control (IS3C). hlm.399–403. Tersedia di <https://ieeexplore.ieee.org/document/9394067> [Accessed 5 Juni 2024].
- Cornia, M., Stefanini, M., Baraldi, L. & Cucchiara, R. 2020. Meshed-Memory Transformer for Image Captioning. *2020 IEEE/CVF Conference on Computer*

- Vision and Pattern Recognition (CVPR)*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). hlm.10575–10584. Tersedia di <https://ieeexplore.ieee.org/document/9157222> [Accessed 5 Juni 2024].
- Deorukhkar, K. & Ket, S. 2022. A detailed review of prevailing image captioning methods using deep learning techniques. *Multimedia Tools and Applications*, 81(1): 1313–1336.
- Dharsini, S.V., Razak, M.A., Modi, S., Reddy, P.K. & Bhatnagar, S. 2022. Captioning Based Image Using Euclidean Distance and ResNet-50. *2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAIAI)*. 2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAIAI). hlm.1–5. Tersedia di <https://ieeexplore.ieee.org/document/10028946> [Accessed 5 Juni 2024].
- Ghandi, T., Pourreza, H. & Mahyar, H. 2024. Deep Learning Approaches on Image Captioning: A Review. *ACM Computing Surveys*, 56(3): 1–39.
- Ghannay, S., Favre, B., Estève, Y. & Camelin, N. 2016. Word Embedding Evaluation and Combination. N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk & S. Piperidis, ed., *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. LREC 2016. Portorož, Slovenia: European Language Resources Association (ELRA), hlm.300–305. Tersedia di <https://aclanthology.org/L16-1046> [Accessed 5 Juni 2024].
- Ghiasi, G., Lin, T.-Y. & Le, Q.V. 2018. *DropBlock: A regularization method for convolutional networks*. Tersedia di <http://arxiv.org/abs/1810.12890> [Accessed 5 Juni 2024].
- Gholamalinezhad, H. & Khosravi, H. 2020. *Pooling Methods in Deep Neural Networks, a Review*. Tersedia di <http://arxiv.org/abs/2009.07485> [Accessed 5 Juni 2024].
- Gillioz, A., Casas, J., Mugellini, E. & Khaled, O.A. 2020. Overview of the Transformer-based Models for NLP Tasks. *Annals of Computer Science and Information Systems*. Proceedings of the 2020 Federated Conference on Computer Science and Information Systems. hlm.179–183. Tersedia di <https://annals-csis.org/proceedings/2020/drp/20.html> [Accessed 5 Juni 2024].

- He, K., Zhang, X., Ren, S. & Sun, J. 2015. *Deep Residual Learning for Image Recognition*. Tersedia di <http://arxiv.org/abs/1512.03385> [Accessed 5 Juni 2024].
- Hochreiter, S. 1998. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06(02): 107–116.
- Hoxha, G., Melgani, F. & Slaghenauffi, J. 2020. A New CNN-RNN Framework For Remote Sensing Image Captioning. *2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*. 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS). hlm.1–4. Tersedia di <https://ieeexplore.ieee.org/document/9105191> [Accessed 6 Juni 2024].
- Koech, K.E. 2024. *How Neural Networks Actually Work — Python Implementation Part 2 (Simplified)*. Medium. Tersedia di <https://towardsdatascience.com/how-neural-networks-actually-work-python-implementation-part-2-simplified-80db0351db45> [Accessed 5 Juni 2024].
- Koto, F., Rahimi, A., Lau, J.H. & Baldwin, T. 2020. *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*. Tersedia di <http://arxiv.org/abs/2011.00677> [Accessed 5 Juni 2024].
- Kumar, G., Kumar, P. & Kumar, D. 2021. Brain Tumor Detection Using Convolutional Neural Network. *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)*. 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC). hlm.1–6. Tersedia di <https://ieeexplore.ieee.org/document/9688460> [Accessed 5 Juni 2024].
- Lin, Z.-X., Qu, J.-X., Yang, X.-B., Hou, Z.-Q. & Ding, C. 2021. Research on Image Caption Based on Multiple Word Embedding Representations. *2021 3rd International Conference on Natural Language Processing (ICNLP)*. 2021 3rd International Conference on Natural Language Processing (ICNLP). hlm.262–266. Tersedia di <https://ieeexplore.ieee.org/document/9537875> [Accessed 5 Juni 2024].
- Liu, J., Lin, X. & He, L. 2023. Image Alone Are Not Enough: A General Semantic-Augmented Transformer-Based Framework for Image Captioning.

- 2023 International Joint Conference on Neural Networks (IJCNN)*. 2023 International Joint Conference on Neural Networks (IJCNN). hlm.01–08. Tersedia di <https://ieeexplore.ieee.org/document/10191656> [Accessed 5 Juni 2024].
- Liu, Z., Dou, Y., Jiang, J. & Xu, J. 2016. Automatic code generation of convolutional neural networks in FPGA implementation. *2016 International Conference on Field-Programmable Technology (FPT)*. 2016 International Conference on Field-Programmable Technology (FPT). hlm.61–68. Tersedia di <https://ieeexplore.ieee.org/document/7929190> [Accessed 5 Juni 2024].
- Lyu, H. 2023. Research on Corrosion Recognition Method of Steel Based on Convolutional Neural Network. *2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE)*. 2023 IEEE 6th International Conference on Information Systems and Computer Aided Education (ICISCAE). hlm.507–511. Tersedia di <https://ieeexplore.ieee.org/document/10393077> [Accessed 5 Juni 2024].
- Mahadi, M.R.S., Arifianto, A. & Ramadhani, K.N. 2020. Adaptive Attention Generation for Indonesian Image Captioning. *2020 8th International Conference on Information and Communication Technology (ICoICT)*. 2020 8th International Conference on Information and Communication Technology (ICoICT). hlm.1–6. Tersedia di <https://ieeexplore.ieee.org/document/9166244> [Accessed 5 Juni 2024].
- Makarencov, V., Shapira, B. & Rokach, L. 2017. *Language Models with Pre-Trained (GloVe) Word Embeddings*. Tersedia di <http://arxiv.org/abs/1610.03759> [Accessed 5 Juni 2024].
- Mao, L. 2019. *Layer Normalization Explained*. Lei Mao's Log Book. Tersedia di <https://leimao.github.io/blog/Layer-Normalization/> [Accessed 5 Juni 2024].
- Moral, Ö.T., Kiliç, V., Onan, A. & Wang, W. 2022. Automated Image Captioning with Multi-layer Gated Recurrent Unit. *2022 30th European Signal Processing Conference (EUSIPCO)*. 2022 30th European Signal Processing Conference (EUSIPCO). hlm.1160–1164. Tersedia di <https://ieeexplore.ieee.org/document/9909888> [Accessed 5 Juni 2024].
- Mulyawan, R., Sunyoto, A. & Muhammad, A.H. 2022. Automatic Indonesian Image Captioning using CNN and Transformer-Based Model Approach. *2022 5th*

International Conference on Information and Communications Technology (ICOIACT). 2022 5th International Conference on Information and Communications Technology (ICOIACT). hlm.355–360. Tersedia di <https://ieeexplore.ieee.org/document/9971855> [Accessed 5 Juni 2024].

Nugraha, A.A., Arifianto, A., & Suyanto 2019. Generating Image Description on Indonesian Language using Convolutional Neural Network and Gated Recurrent Unit. *2019 7th International Conference on Information and Communication Technology (ICoICT)*. 2019 7th International Conference on Information and Communication Technology (ICoICT). hlm.1–6. Tersedia di <https://ieeexplore.ieee.org/document/8835370> [Accessed 5 Juni 2024].

Panigrahi, L., Panigrahi, R.R. & Chandra, S.K. 2023. Hybrid Image Captioning Model. *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*. 2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON). hlm.1–6. Tersedia di <https://ieeexplore.ieee.org/document/10113957> [Accessed 5 Juni 2024].

Papineni, K., Roukos, S., Ward, T. & Zhu, W.-J. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. P. Isabelle, E. Charniak & D. Lin, ed., *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. ACL 2002. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, hlm.311–318. Tersedia di <https://aclanthology.org/P02-1040> [Accessed 5 Juni 2024].

Pere, C. 2020. *What are Loss Functions?* Medium. Tersedia di <https://towardsdatascience.com/what-is-loss-function-1e2605aeb904> [Accessed 5 Juni 2024].

Phuong, M. & Hutter, M. 2022. *Formal Algorithms for Transformers*. Tersedia di <http://arxiv.org/abs/2207.09238> [Accessed 5 Juni 2024].

Raiaan, M.A.K., Sakib, S., Fahad, N.M., Mamun, A.A., Rahman, Md.A., Shatabda, S. & Mukta, Md.S.H. 2024. A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks. *Decision Analytics Journal*, 11: 100470.

Raschka, S. *Understanding and Coding the Self-Attention Mechanism of Large Language Models From Scratch*. Sebastian Raschka, PhD. Tersedia di

<https://sebastianraschka.com/blog/2023/self-attention-from-scratch.html>

[Accessed 5 Juni 2024].

Segarceanu, S. & Suci, G. 2022. Convolutional Neural Networks for Environmental Sound Recognition. *2022 21st RoEduNet Conference: Networking in Education and Research (RoEduNet)*. 2022 21st RoEduNet Conference: Networking in Education and Research (RoEduNet). hlm.1–5. Tersedia di <https://ieeexplore.ieee.org/document/9920993> [Accessed 5 Juni 2024].

Sharma, D., Dingliwal, R., Dhiman, C. & Kumar, D. 2022. Lightweight Transformer with GRU Integrated Decoder for Image Captioning. *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). hlm.434–438. Tersedia di <https://ieeexplore.ieee.org/document/10090020> [Accessed 5 Juni 2024].

Sharma, S., Sharma, S. & Athaiya, A. 2020. ACTIVATION FUNCTIONS IN NEURAL NETWORKS. *International Journal of Engineering Applied Sciences and Technology*, 04(12): 310–316.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56): 1929–1958.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. 2015. *Rethinking the Inception Architecture for Computer Vision*. Tersedia di <http://arxiv.org/abs/1512.00567> [Accessed 5 Juni 2024].

Tsang, S.-H. 2019. Review: Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015. *Medium*. Tersedia di <https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c> [Accessed 5 Juni 2024].

Tun, N.L., Gavrilov, A. & Tun, N.M. 2020. Multi-Classification of Satellite Imagery Using Fully Convolutional Neural Network. *2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*. 2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). hlm.1–5. Tersedia di <https://ieeexplore.ieee.org/document/9111928> [Accessed 5 Juni 2024].

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser,

- L. & Polosukhin, I. 2023. *Attention Is All You Need*. Tersedia di <http://arxiv.org/abs/1706.03762> [Accessed 5 Juni 2024].
- Voditel, P., Gurjar, A., Pandey, A., Jain, A., Sharma, N. & Dubey, N. 2023. Image Captioning - A Deep Learning Approach Using CNN and LSTM Network. *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*. 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN). hlm.343–348. Tersedia di <https://ieeexplore.ieee.org/document/10265997> [Accessed 5 Juni 2024].
- Wang, S., Li, B.Z., Khabsa, M., Fang, H. & Ma, H. 2020. *Linformer: Self-Attention with Linear Complexity*. Tersedia di <http://arxiv.org/abs/2006.04768> [Accessed 5 Juni 2024].
- Zheng, J., Ma, L., Wu, Y., Ye, L. & Shen, F. 2022. Nonlinear Dynamic Soft Sensor Development with a Supervised Hybrid CNN-LSTM Network for Industrial Processes. *ACS Omega*, 7(19): 16653–16664.