

## DAFTAR PUSTAKA

- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., & Nahavandi, S. (2021). A review of uncertainty quantification in *deep learning*: Techniques, applications and challenges. *Information Fusion*, 76, 243–297. <https://doi.org/10.1016/j.inffus.2021.05.008>
- Banerjee, S., Lavie, A., & Association for Computational Linguistics. (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* (pp. 65–72). <https://aclanthology.org/W05-0909.pdf>
- Cai, C., Yap, K., & Wang, S. (2022). Attribute conditioned *fashion Image captioning*. *2022 IEEE International Conference on Image Processing (ICIP)*, 1921–1925. <https://doi.org/10.1109/icip46576.2022.9897417>
- Carter, R. (2024, September 26). Writing Incredible Product *Descriptions* for Clothes: The Ultimate Guide for Retailers. *NovaTomato*. [novatomato.com/blog/writing-incredible-product-descriptions-for-clothes](https://novatomato.com/blog/writing-incredible-product-descriptions-for-clothes)
- Cornia, M., Stefanini, M., Baraldi, L., & Cucchiara, R. (2020). Meshed-Memory *transformer* for *image captioning*. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10575–10584. <https://doi.org/10.1109/cvpr42600.2020.01059>
- Del Moro, M., Tudosie, S. C., Vannoni, F., Galassi, A., & Ruggeri, F. (2023). Inception Models for *fashion Image captioning*: An extensive study on multiple datasets. In *Lecture notes in computer science* (pp. 3–14). [https://doi.org/10.1007/978-3-031-42448-9\\_1](https://doi.org/10.1007/978-3-031-42448-9_1)
- Dwivedi, P., & Upadhyaya, A. (2022). A novel *deep learning* model for accurate prediction of image captions in *fashion* industry. *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. <https://doi.org/10.1109/confluence52989.2022.9734171>
- Gamini, D., & Perera, K. (2024, July 17). *Empowering the Captioning of Fashion Attributes from Asian Fashion Images*. <https://ijrcom.org/index.php/ijrc/article/view/131>
- Hacheme, G., & Sayouti, N. (2021). Neural *fashion Image captioning*: Accounting for data Diversity. <https://doi.org/10.48550/arxiv.2106.12154>
- Han, K., Guo, J., Zhang, C., & Zhu, M. (2018). Attribute-Aware *attention* model for fine-grained representation learning. *Proceedings of the 30th ACM International Conference on Multimedia*. <https://doi.org/10.1145/3240508.3240550>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/cvpr.2016.90>
- Herdade, S., Kappeler, A., Boakye, K., & Soares, J. (n.d.). *Image Captioning: Transforming Objects into Words*. Retrieved June 6, 2025, from <https://dl.acm.org/doi/pdf/10.5555/3454287.3455286>

- Kong, S., & Fowlkes, C. (2017). Low-Rank Bilinear Pooling for Fine-Grained Classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7025–7034. <https://doi.org/10.1109/cvpr.2017.743>
- Madison, H. (2023, June 2). *The importance of detailed product descriptions for e-commerce stores*. Fibre2Fashion. Retrieved April 10, 2025, from <https://www.fibre2fashion.com/industry-article/9658/the-importance-of-detailed-product-descriptions-for-e-commerce-stores>
- Moratelli, N., Barraco, M., Morelli, D., Cornia, M., Baraldi, L., & Cucchiara, R. (2023). *Fashion-Oriented Image Captioning with External Knowledge Retrieval and Fully Attentive Gates*. *Sensors*, 23(3), 1286. <https://doi.org/10.3390/s23031286>
- Nguyen, B. T., Prakash, O., & Vo, A. H. (2020). *Attention mechanism for fashion image captioning*. In *Advances in intelligent systems and computing* (pp. 93–104). [https://doi.org/10.1007/978-3-030-62324-1\\_9](https://doi.org/10.1007/978-3-030-62324-1_9)
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2001). *BLEU*. *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 311. <https://doi.org/10.3115/1073083.1073135>
- Payments & Commerce Market Intelligence. *The E-Commerce Data Portrait of Asia Pacific*. (2024). [https://americasmi.com/pdfs\\_landings/PCMI\\_Asia\\_Ecommerce\\_Data\\_Por](https://americasmi.com/pdfs_landings/PCMI_Asia_Ecommerce_Data_Por)
- Shetty, A., Kale, Y., Patil, Y. dkk. *Optimal transformers based image captioning using beam search*. *Multimed Tools Appl* 83, 47963–47977 (2024). <https://doi.org/10.1007/s11042-023-17359-6trait.pdf>
- Shopify. (2023). *SEO Product Descriptions: 7 Tips To Optimize Your Product Pages*. Retrieved from <https://www.shopify.com/enterprise/blog/seo-product-descriptions>
- Standard Insights. (2023, May 10). *eCommerce in Indonesia: Shopper trends & top sites*. <https://standard-insights.com/insights/the-rise-of-ecommerce-in-indonesia/>
- Tang, Y., Zhang, L., & Yuan, Y. (2023). *Fashion item captioning via grid-relation self-attention and gated-enhanced decoder*. *Multimedia Tools and Applications*, 83(3), 7631–7655. <https://doi.org/10.1007/s11042-023-15492-w>
- Tang, Y., Zhang, L., Yuan, Y., & Chen, Z. (2023). *Improving fashion captioning via attribute-based alignment and multi-level language model*. *Applied Intelligence*, 53(24), 30803–30821. <https://doi.org/10.1007/s10489-023-05167-2>
- Tiwari, V., Joshi, R. C., & Dutta, M. K. (2022). *Deep neural network for multi-class classification of medicinal plant leaves*. *Expert Systems*, 39(8). <https://doi.org/10.1111/exsy.13041>
- Tu, G., Liu, Y., & Vlassov, V. (2023). *AIC-AB NET: A Neural Network for Image Captioning with Spatial Attention and Text Attributes*. <https://doi.org/10.48550/arxiv.2307.07370>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*.

- Vedantam, R., Zitnick, C. L., & Parikh, D. (2015). *CIDEr: Consensus-based image description evaluation*. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4566–4575. <https://doi.org/10.1109/cvpr.2015.7299087>
- Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. 3156–3164. <https://doi.org/10.1109/CVPR.2015.7298935>
- Yu, J., Li, J., Yu, Z., & Huang, Q. (2020). Multimodal transformer with multi-view visual representation for image captioning. *IEEE Trans. Circuits and Systems for Video Technol.*, 30, 12. <https://doi.org/10.1109/TCSVT.2019.2947482>