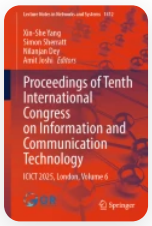
Enhanced Aerial Scene Classification Through ConvNeXt Architectures and Channel Attention

| Conference paper | First Online: 01 October 2025

| pp 225–236 | Cite this conference paper



Proceedings of Tenth International Congress on Information and Communication Technology

(ICICT 2025)

Leo Thomas Ramos & Angel D. Sappa

Part of the book series: Lecture Notes in Networks and Systems ((LNNS,volume 1412))

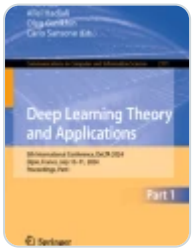
Included in the following conference series:
International Congress on Information and Communication Technology

Abstract

This work explores the integration of a Channel Attention (CA) module into the ConvNeXt architecture to improve performance in scene classification tasks. Using the UC Merced

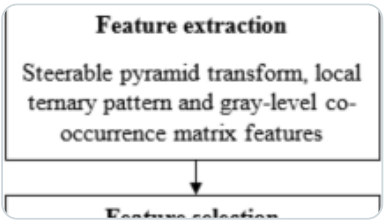
dataset, experiments were conducted with two data splits: 50% and 20% for training. Models were trained for up to 20 epochs, limiting the training process to assess which models could extract the most relevant features efficiently under constrained conditions. The ConvNeXt architecture was modified by incorporating a Squeeze-and-Excitation block, aiming to enhance the importance of each feature channel. ConvNeXt models with CA showed strong results, achieving the highest performance in the experiments conducted. ConvNeXt large with CA reached 90% accuracy and 89.75% F1-score with 50% of the training data, while ConvNeXt base with CA achieved 77.14% accuracy and 75.23% F1-score when trained with only 20% of the data. These models consistently outperformed their standard counterparts, as well as other architectures like ResNet and Swin Transformer, achieving improvements of up to 9.60% in accuracy, highlighting the effectiveness of CA in boosting performance, particularly in scenarios with limited data.

Similar content being viewed by others



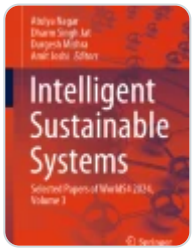
Evolving Deep Architectures: A New Blend of CNNs and Transformers Without…

Chapter | © 2024



UFS-LSTM: unsupervised feature selection with long short-term memory…

Article | 30 August 2021



Scene and Aerial Image Classification Based on Advanced Deep Learning Technique

Chapter | © 2025

Notes

1. http://weegee.vision.ucmerced.edu/datasets/landuse.html.

2. https://pytorch.org/vision/stable/models.html.

References

1. Chen SB, Wei QS, Wang WZ, Tang J, Luo B, Wang ZY (2022) Remote sensing scene classification via multi-branch local attention network. IEEE Trans Image Process 31:99–109. https://doi.org/10.1109/TIP.2021.3127851

Article Google Scholar

2. Cheng G, Xie X, Han J, Guo L, Xia GS (2020) Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. IEEE J Sel Top Appl Earth Obs Rem Sens 13:3735–3756. https://doi.org/10.1109/JSTARS.2020.3005403

Article Google Scholar

3. Guo N, Jiang M, Wang D, Zhou X, Song Z, Li Y, Gao L, Luo J (2024) Scene classification for remote sensing image of land use and land cover using dual-model architecture

with multilevel feature fusion. Int J Digit Earth 17(1). https://doi.org/10.1080/17538947.2024.2353166

4. Hu J, Shen L, Albanie S, Sun G, Wu E (2020) Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 42(8):2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

Article Google Scholar

5. Kim B, Yuvaraj N, Sri Preethaa KR, Arun Pandian R (2021) Surface crack detection using deep learning with shallow cnn architecture for enhanced computation. Neural Comput Appl 33(15):9289–9305. https://doi.org/10.1007/s00521-021-05690-8

Article Google Scholar

6. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S (2022) A convnet for the 2020s. In: 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 11966–11976. https://doi.org/10.1109/CVPR52688.2022.01167

7. Ma A, Wan Y, Zhong Y, Wang J, Zhang L (2021) Scenenet: Remote sensing scene classification deep learning network using multi-objective neural evolution architecture search. ISPRS J Photogram Rem Sens 172:171–188. https://doi.org/10.1016/j.isprsjprs.2020.11.025

Article Google Scholar

8. Mei S, Yan K, Ma M, Chen X, Zhang S, Du Q (2021) Remote sensing scene classification using sparse representation-based framework with deep feature fusion. IEEE J Select Top Appl Earth Obs Rem Sens 14:5867–5878. https://doi.org/10.1109/JSTARS.2021.3084441

Article Google Scholar

9. Niu Z, Zhong G, Yu H (2021) A review on the attention mechanism of deep learning. Neurocomputing 452:48–62. https://doi.org/10.1016/j.neucom.2021.03.091

Article Google Scholar

10. Ramos LT, Sappa AD (2024) Multispectral semantic segmentation for land cover classification: An overview. IEEE J Select Top Appl Earth Obs Rem Sens 17:14295–14336. https://doi.org/10.1109/JSTARS.2024.3438620

Article Google Scholar

11. Tang X, Ma Q, Zhang X, Liu F, Ma J, Jiao L (2021) Attention consistent network for remote sensing scene classification. IEEE J Select Top Appl Earth Obs Rem Sens 14:2030–2045. https://doi.org/10.1109/JSTARS.2021.3051569

Article Google Scholar

12. Tong W, Chen W, Han W, Li X, Wang L (2020) Channel-attention-based densenet network for remote sensing image scene classification. IEEE J Select Top Appl Earth Obs Rem Sens 13:4121–4132. https://doi.org/10.1109/JSTARS.2020.3009352

Article Google Scholar

13. Wang J, Li W, Zhang M, Tao R, Chanussot J (2023) Remote-sensing scene classification via multistage self-guided separation network. IEEE Trans Geosci Rem Sens 61:1–12. https://doi.org/10.1109/TGRS.2023.3295797

Article Google Scholar

14. Zhu C, Ping W, Xiao C, Shoeybi M, Goldstein T, Anandkumar A, Catanzaro B (2021) Long-short transformer: Efficient transformers for language and vision. In: Advances in neural information processing systems, vol 34. Curran Associates, Inc., pp 17723–17736

Google Scholar

Acknowledgements

This was supported in part by Grant PID2021-128945NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". The authors acknowledge the support of the Generalitat de Catalunya CERCA Program to CVC's general activities, and the Departament de Recerca i Universitats from Generalitat de Catalunya with reference 2021SGR 01499.

Author information

Authors and Affiliations

Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain
Leo Thomas Ramos & Angel D. Sappa

Kauel Inc., Menlo Park, USA
Leo Thomas Ramos

ESPOL Polytechnic University, Guayaquil, Ecuador
Angel D. Sappa

Corresponding author

Correspondence to Leo Thomas Ramos .

Editor information

Editors and Affiliations

Department of Design Engineering and Mathematics, Middlesex University London, London, UK
Xin-She Yang

Department of Biomedical Engineering, The University of Reading, Berkshire, UK
Simon Sherratt

Department of Computer Science and Engineering, Techno International New Town, Kolkata, West Bengal, India
Nilanjan Dey

Global Knowledge Research Foundation, Ahmedabad, Gujarat, India
Amit Joshi

Rights and permissions

Copyright information

About this paper

Cite this paper

Ramos, L.T., Sappa, A.D. (2025). Enhanced Aerial Scene Classification Through ConvNeXt Architectures and Channel Attention. In: Yang, XS., Sherratt, S., Dey, N., Joshi, A. (eds) Proceedings of Tenth International Congress on Information and Communication Technology. ICICT 2025. Lecture Notes in Networks and Systems, vol 1412. Springer, Singapore. https://doi.org/10.1007/978-981-96-6429-0_19

.RIS .ENW .BIB

DOI	Published	Publisher Name
https://doi.org/10.1007/978-981-96-6429-0_19	01 October 2025	Springer, Singapore

Print ISBN	Online ISBN	eBook Packages
978-981-96-6428-3	978-981-96-6429-0	Engineering Engineering (R0)

Keywords

Scene classification Remote sensing Convolutional neural networks

Attention mechanisms Computer vision

Publish with us

Policies and ethics