



Universidad
Internacional
de Valencia

Uso de modelos de visión por computador basados en Transformers en tiempo real en dispositivos móviles

Titulación:
Máster en Inteligencia
Artificial
Curso académico
2023 – 2024

Alumno/a: La Casa Nieto,
José Jesús
D.N.I.: 77249230G

Director/a de TFM: Enrique
Mas Candela

Convocatoria:
Segunda convocatoria

Dedicatoria

Escribir dedicatoria aquí.

Agradecimientos

Escribir agradecimientos aquí

Agradecimientos.

Resumen

Abstract

Índice general

Dedicatoria	I
Agradecimientos	III
Resumen	V
Abstract	VII
Índice general	IX
Índice de figuras	XI
Índice de tablas	XIII
1. Introducción	1
2. Estado del arte	3
3. Marco teórico	5
4. Materiales	7
5. Métodos	9
6. Resultados	11
Bibliografía	13

Índice de figuras

Índice de tablas

“Frase.” - Autor

Capítulo 1

Introducción

Capítulo 2

Estado del arte

Capítulo 3

Marco teórico

Capítulo 4

Materiales

Capítulo 5

Métodos

Capítulo 6

Resultados

Bibliografía

- [1] Pan, J.; Bulat, A.; Tan, F.; Zhu, X.; Dudziak, L.; Li, H.; Tzimiropoulos, G.; Martinez, B. (2022). *EdgeViTs: Competing Light-weight CNNs on Mobile Devices with Vision Transformers*. The Chinese University of Hong Kong, Samsung AI Cambridge, Queen Mary University of London.
- [2] Li, Y.; Yuan, G.; Wen, Y.; Hu, J.; Evangelidis, G.; Tulyakov, S.; Wang, Y.; Ren, J. (2022). *EfficientFormer: Vision Transformers at MobileNet Speed*. Snap Inc, Northeastern University.
- [3] Chen, Y.; Dai, X.; Chen, D.; Liu, M.; Dong, X.; Yuan, L.; Liu, Z. (2022). *Mobile-Former: Bridging MobileNet and Transformer*. Microsoft, University of Science and Technology of China.
- [4] Mehta, S.; Rastegari, M. (2022). *MobileViT: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer*. Apple.
- [5] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.; Kaiser, L.; Polosukhin, I. (2017). *Attention Is All You Need*. Google, University of Toronto.
- [6] Radford, A.; Kim, J.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision*. OpenAI.
- [7] Ma, S.; Wang, H.; Ma, L.; Wang, L.; Wang, W.; Huang, S.; Dong, L.; Wang, R.; Xue, J.; Wei, F. (2024). *The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits*. Microsoft.
- [8] Howard, A.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Network for Mobile Vision Applications*. Google.

- [9] Wadekar, S.; Chaurasia, A. (2022). *MobileViT3: Mobile-Friendly Vision Transformer With Simple and Effective Fusion of Local, Global and Input Features*. Micron Technology.
- [10] Howard, A.; Sandler, M.; Chu, G.; Chen, L.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; Le, Q.; Adam, H. (2019). *Searching for MobileNetV3*. Google AI, Google Brain.
- [11] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; Houlsby, N. (2020). *An Image is Worth 16x16: Transformers for Image Recognition at Scale*. Google Research, Brain Team.
- [12] Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.; Shahbaz, F.; Shah, M. (2021). *Transformers in Vision: A Survey*. ACM Computing Surveys (CSUR).
- [13] Abdelhamed, A.; Lin, S.; Brown, M. (2018). *A High-Quality Denoising Dataset for Smartphone Cameras*. York University, Microsoft.
- [14] Abdelhamed, A.; Timofte, R.; Brown, M.; Yu, S.; Park, B.; Jeong, J.; Jung, S. (2019). *NTIRE 2019 Challenge on Real Image Denoising: Methods and Results*. Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).
- [15] Nah, S.; Kim, T.; Lee, K. (2016). *Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring*. Seoul National University.
- [16] Conde, M.; Vasluianu, F.; Vazquez-Corral, J.; Timofte, R. (2022). *Perceptual Image Enhancement for Smartphone Real-Time Applications*. University of Würzburg, Universitat Autònoma de Barcelona.
- [17] Feng, R.; Li, C.; Chen, H.; Li, S.; Loy, C.; Gu, J. (2021). *Removing Diffraction Image Artifacts in Under-Display Camera via Dynamic Skip Connection Network*. Nanyang Technological University, Tetras AI, Shanghai AI Laboratory.
- [18] Ekman, M. (2022). *Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, Natural Language Processing, and Transformers Using TensorFlow*. Boston: Addison-Wesley.
- [19] Lakshmanan, V.; Görner, M.; Gillard, R. (2021). *Practical Machine Learning for Computer Vision*. Beijing: O'Reilly.

- [20] Rothman, D. (2024). *Transformers for Natural Language Processing and Computer Vision*. Birmingham-Mumbai: Packt.



Universidad
Internacional
de Valencia

14 Octubre 2024

De:
 Planeta Formación y Universidades