Rubèn Pérez Tito

rperez@cvc.uab.cat
Computer Vision Center, Campus UAB, Barcelona
+34 627 599 005

I have recently completed my Ph.D. on "Exploring the role of Text in Visual Question Answering on Natural Scenes and Documents". Highlighting the importance of text in natural scenes, and using VQA as a natural language interface to guide information extraction on document images. During this journey, I analyzed the needs and flaws of existing research and industry, pursuing the most interesting topics such as integrating and exploiting new modalities in VQA, or processing long sequences of tokens. Moreover, I've always been a very versatile person, being able to explore new challenges and domains either because they were interesting or to help some colleagues in their research. Thanks to this, although I started my research working on scene-text image retrieval and handwritten documents, I moved soon to the challenging and multimodal VQA task, which allowed me to investigate and pursue my Ph.D. During which I have also worked on model calibration, federated learning or privacy preservation with differential privacy.
I enjoy participating in Artificial Intelligence panels and workshops open to the public at Barcelona because one of my aims is to bring the progress and concerns from research centers to the public. Beyond my scientific profile, I love to go to the mountains to ease my mind and admire the beauty of the nature. On the other hand, I enjoy participating in Artificial Intelligence panels open to the public at Barcelona because one of my aims is to bring the progress and concerns from the research centers to the public.

Education

2018 – 2023 **Ph.D. in Computer Science, Universitat Autònoma de Barcelona, Spain**
Advisor: Dr. Ernest Valveny
Thesis title: *Exploring the role of Text in Visual Question Answering on Natural Scenes and Documents.*

2016 – 2018 **M.Sc. in Computer Engineering, Universitat Autònoma de Barcelona**
Advisor: Dr. Antonio Espinosa
Thesis title: *Acceleration of Big Data iterative workflows with MapD-GPU.*

2010 – 2016 **B.Sc. in Computer Engineering, Universitat Autònoma de Barcelona**

Research/Employment Experience

2021 – 2023 **Teaching assistant.**
Master in Computer Vision, Universitat Autónoma de Barcelona
Subject: Object detection, recognition, image retrieval and cross-modal retrieval.

2018 – 2021 **Teaching assistant.**
Computer Science Engineering, Universitat Autónoma de Barcelona
Subject: Fundamentals of programming in C.

Research/Employment Experience (continued)

2018 **Intern.**
Computer Vision Center, Universitat Autónoma de Barcelona
Subject: Single shot scene text retrieval. **Advisor: Dr. Dimosthenis Karatzas**

2015 – 2016 **Junior developer.**
Banking and Finance consulting department, Everis Spain S.L.U.
Subject: Bank solutions development with IBM Business Process Manager.

Publications

Most of my contributions have focused on seeking for new tasks that were useful for the society and industry, focusing on my Ph.D. topic VQA. In this regard, we first proposed a dataset [12, 11] and model [6] that integrated scene text present in natural images to provide literate abilities to VQA models. Then, we used the VQA paradigm as a natural language interface to extract information from documents, which allowed to agglutinate several different tasks into a single one. We initially start with gray-scale industry documents [10] and then, moved to more visually complex infographic images [5]. Afterward, we changed the perspective and instead of using single images. We focused on extracting information from multiple document images at the same time. First, document collections of independent single page documents [8], and multipage documents [1] afterward. Moreover, in [2, 3] we also included model calibration in VQA, to assess if the models' confidence are aligned with their performance, and incorporating the challenging non-answerable questions. In the last work, we explore privacy preservation in document visual question answering in distributed federated learning scenarios, which focuses on alleviating current privacy concerns.

1. **Tito, Rubèn**, Karatzas, D. & Valveny, E. (2023). Hierarchical multimodal transformers for multipage docvqa. *Pattern Recognition*, *144*, 109834.
2. Van Landeghem, J., **Tito, R.**, Borchmann, Ł., Pietruszka, M., Joziak, P., Powalski, R., … Valveny, E. et al. (2023). Document understanding dataset and evaluation (dude). In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 19528–19540).
3. Van Landeghem, J., Tito, R., Borchmann, Ł., Pietruszka, M., Jurkiewicz, D., Powalski, R., … Stanisławek, T. (2023). Icdar 2023 competition on document understanding of everything (dude). In *International conference on document analysis and recognition* (pp. 420–434). Springer.
4. Biten, A. F., Tito, R., Gomez, L., Valveny, E. & Karatzas, D. (2022). Ocr-idl: Ocr annotations for industry document library dataset. In *European conference on computer vision* (pp. 241–252). Springer.
5. Mathew, M., Bagal, V., **Tito, R.**, Karatzas, D., Valveny, E. & Jawahar, C. (2022). Infographicvqa. In *Proceedings of the ieee/cvf winter conference on applications of computer vision* (pp. 1697–1706).
6. Gómez, L., Biten, A. F., **Tito, R.**, Mafla, A., Rusiñol, M., Valveny, E. & Karatzas, D. (2021). Multimodal grid features and cell pointers for scene text visual question answering. *Pattern Recognition Letters*, *150*, 242–249.
7. Mafla, A., **Tito, R.**, Dey, S., Gomez, L., Rusinol, M., Valveny, E. & Karatzas, D. (2021). Real-time lexicon-free scene text retrieval. *Pattern Recognition*, *110*, 107656.
8. **Tito, R.**, Karatzas, D. & Valveny, E. (2021). Document collection visual question answering. In *2021 international conference on document analysis and recognition (icdar)*. Springer.
9. **Tito, R.**, Mathew, M. & Karatzas, D. (2021). Icdar 2021 competition on document visual question answering. In *International conference on document analysis and recognition* (pp. 635–649). Springer.

10 Mathew, M., **Tito, R.**, Karatzas, D., Manmatha, R. & Jawahar, C. (2020). Document visual question answering challenge 2020. *arXiv preprint arXiv:2008.08899*.

11 *Biten, A. F., ***Tito, Ruben**, *Mafla, A., Gomez, L., Rusinol, M., Mathew, M., … Karatzas, D. (2019). Icdar 2019 competition on scene text visual question answering. In *2019 international conference on document analysis and recognition (icdar)* (pp. 1563–1570). IEEE.

12 *Biten, A. F., ***Tito, R.**, *Mafla, A., Gomez, L., Rusinol, M., Jawahar, C., … Karatzas, D. (2019). Scene text visual question answering. In *Proceedings of the international conference on computer vision (iccv 2019)*.

Skills

Languages	English (Fluent), Spanish (Native), Catalan (Native)
Coding	Python, Java, Scala, C, C++, HTML, CSS, Javascript, SQL
Standards and methodologies	ISO 9000, ISO 27000, BPMN, Scrum, TDD, UML
Frameworks	PyTorch, scikit-learn, MySQL, MongoDB, Unity
Misc.	Science fiction books, Skiing, Cooking, Hiking, Travelling

References

Dr. Ernest Valveny
Computer Vision Centre, Room 122, Edifici O
Universitat Autònoma de Barcelona
☎ +34 93 581 18 28
✉ ernest@cvc.uab.es

Dr. Dimosthenis Karatzas
Computer Vision Centre, Room 124, Edifici O
Universitat Autònoma de Barcelona
☎ +34 93 581 38 41
✉ dimos@cvc.uab.es