João P C Bertoldo

26, Brazilian

Online CV | One-page CV

joaopcbertoldo.github.io

April, 2021

(+33) 06 25 51 62 57

joao.bertoldo@mines-paristech.fr

0000-0002-9512-772X

Curriculum Vitae

Summary

Education

Double degree MSc. in artificial intelligence [a]

Double degree MSc. in engineering [b,d]

Research

1 year of experience [f,g]

Currently working as research engineer [e].

Work (others)

2 years of experience as Software Engineer [i,j], Data Scientist [h,j], and Teaching Assistant [l].

Experience in software development with diverse technologies.

Projects, Skills, etc

Using Python for several years in individual [o,p] and in-collaboration projects [q,r].

Lived in four countries – Brazil [d,j], France [a,b,e,g], Norway [i], USA [h], and speaks three languages.

1. Education

Paris-Dauphine – PSL University [(a)] — Paris, France
MSc. in Artificial Intelligence, Systems, Data (IASD) — 2019 – 2020
Double degree with [b]
Subjects: Machine Learning, Deep Learning, Optimization, Image Analysis

MINES ParisTech – PSL University [(b)] — Paris, France
MSc. in Executive Engineering – Minor: Data Science — 2017 – 2020
Double degree with [d]
Subjects: Software Engineering, Databases, Probability, Statistics, Operations Research

ENSIAME – INSA Hauts-de-France [(c)] — Valenciennes, France
Mechatronics Engineering — 2015 – 2016
1-year scholarship for an international academic exchange programme

University of São Paulo (USP) [(d)] — São Carlos, Brazil
Mechatronics Engineering — 2013 – 2015, 2016 – 2017
5-year degree at the best university in Latin America (U.S. News & World Report Best Colleges Ranking, 2020)

2. Research

The European Synchrotron Radiation Facility (ESRF) [(e)] — Grenoble, France
MINES ParisTech – PSL University — Paris, France
Research Engineer — since March 2021
Working on X-ray diffraction tomography (DCT) semantic segmentation.

MINES ParisTech – PSL University [(f)] — Paris, France
Research Intern — 2020 – 2021 (6 mos.)
Worked on semantic segmentation with deep learning (2D and 3D U-Nets) applied to X-ray tomography (XCT).
Published the code [o], data [1], and models [2].
Presented results in two seminars [m,n], and lectured in a week-course [k].

Mindsay [g] — Paris, France
Research Intern — 2020 (5 mos.)
Evaluated explainability methods layperson-maintained Natural Language Processing (NLP) classification models.
References: LIME (Ribeiro et al., 2016), SHAP (Ribeiro et al., 2018), Anchors (Lundberg et al., 2017)
→ *Links to: internship report, internship defense presentation*

3. Work (others)

Datadog [h] — New York, USA
Data Scientist Intern — 2019 (7 mos.)
Developed algorithms for time series and tag analysis [q], and contributed to a knowledge graph project.
Maintained high-impact production Python codebase and optimized an automated test pipeline.

Kelda Drilling Controls [i] — Porsgrunn, Norway
Software Engineer Intern — 2018 – 2019 (8 mos.)
Designed a web application for time series supervised annotation [r].

Stone Co. [j] — Rio de Janeiro, Brazil
Backend Developer / Data Scientist Intern — 2017 – 2018 (3 mos. / 2 mos.)
Developed a .NET web Rest API for a MongoDB and lectured about Design Patterns.
Explored transactional data to predict churn risk with decision trees-based methods.

4. Teaching

Computer vision and machine learning for the material scientist (CVML2021) [k] — Paris, France
Member of the teaching team — 2021 (1 week)
Lectured about convolutional neural networks applied to semantic segmentation of composite material tomography.
Assisted students during practical sessions.
→ *Link to my slides*

University of São Paulo (USP) [l] — São Carlos, Brazil
Teaching assistant (part-time) — 2015 (6 mos.)
Elaborated and conducted workshops on numerical analysis using MATLAB.

5. Talks

44th edition of the Journée ISS France [m] — Online
Deep learning for automated segmentation of tomographic images — Feb. 2021
→ *Link to my slides*

Seminar at the Center for Mathematical Morphology (CMM) [n] — Online
MINES ParisTech – PSL University — Jan. 2021
Fiber composite 3D segmentation with neural networks
→ *Link to my slides*

6. Projects

tomo2seg [o] — since Sep. 2020 [f,e]
Developed an X-ray tomography (XCT) semantic segmentation tool using U-Nets.
Applied it during an experiment at the Soleil Synchrotron , France.
Data [1], models [2], and a complementary tutorial (WIP) are publicly available.
→ Link to tomo2seg on GitHub

pymdr [p] — 2020 (1 mo.) [a]
Python implementation of *Mining Data Records* (Liu et al., 2003)

→ Link to pymdr on GitHub

Correlations [q] 2019 (7 mos.) [h]

Contributed to build *Correlations*: an automated debug tool for complex infrastructures capable of finding correlated accidents in time series.

→ Link to Correlations on Datadog's website

Bivrost [r] 2018 – 2019 (8 mos.) [i]

Maintained and implemented new features in *Bivrost*: a web application for high-resolution time series visualization used for real-time drilling systems monitoring.

Designed an integrated time series annotation tool for supervised machine learning.

→ Link to Bivrost's page on Kelda Drilling Controls's website

Min'light: sunlight simulator [s] 2017 – 2018 (9 mos.) [b]

Led a project for 6 mos. to build a robotic cable-controlled sunlight physical simulator.

Developed a 3D motion digital twin in Python.

→ Link to *minlight* on GitHub

MKafeina: web-controlled coffee machine [t] 2016 (4 mos.) [d]

Designed a coffee machine controlled over the internet.

Developed a web server in .NET (C#) and an Arduino controller (C++).

→ Link to *mkafeina* on GitHub

7. Skills

Languages

English Fluent

Obtained TOEFL 117/120 (2020) and worked in English-speaking companies for more than 1 year [h,i].

Wrote Master's end-of-studies internship report in English [g].

French Fluent

Read subjects in French for 3 years [a,b,c] and worked in French organizations for 1 year [e,f,g].

Lectured in French [m,n,k].

Portuguese Native

Turkish Beginner

Programming

Python Fluent

Data processing/viz. Numpy, Scipy, Pandas, IPython, Matplotlib, Seaborn

Machine/Deep learning Keras/TensorFlow [o], Scikit-learn [j]

Prototyping Streamlit [g], Jupyter Lab/Notebook [f,g,h]

Miscellaneous Anaconda, virtualenv, nose, unittests

Other programming languages

MATLAB [l], JavaScript (React, Redux) [i], C/C++ (basics) [d,t], Java (basics) [b]

Databases

PostgresSQL [b,j], MongoDB [i,j], InfluxDB [i]

Computer skills

IDEs Pycharm, Visual Studio Code (VSCode)

Tomography ImageJ (Fiji), Avizo

Miscellaneous Git, Linux, Bash/Zsh, SSH, Markdown, Weka

Content creation/management

LaTeX Fluent.

Wrote Master's end-of-studies internship report with LaTeX [g].

Google Suite (Docs, Slides, Sheets)

Presentations prepared with Google Slides: [k], [m], [n], [g].

Zotero Extensively using for more than 1 year.

8. Interests

Academic

Deep learning

Attention Is All You Need (Vaswani et al., 2017)

The Lottery Ticket Hypothesis (Frankle and Carbin, 2019)

Computer Vision, AI applied to the Sciences and Art, 3D X-ray tomography

Newsletters

Papers with Code, DeepAI, The Variable, The Batch (DeepLearning.AI)

Miscellaneous

Papers with Code, Papers *without* Code, Distill, Towards Data Science

Personal

Travel & international environments

Have been in 18 countries.

Skydiving

Member of a skydiving association (ASPU)

Podcasts

(PT) Cafe da manhã, Foro de Teresina

(EN) Do you really know?, Impact

(FR) Maintenant vous savez, Débat du jour, Le moment Meurice, Le monde devant soi, Les couilles sur la table, Madame meuf, Programme B

Youtube channels

Two Minute Papers, 3Blue1Brown, Vox, The Economist, Porta dos Fundos, Choque de Cultura

Publications

[1] J. Bertoldo, E. Decencière, D. Ryckelynck, and H. Proudhon, *Glass fiber-reinforced polyamide 66 3D X-ray computed tomography dataset for deep learning segmentation*, eng, Mar. 2021. DOI: 10.5281/zenodo.4587827.

[2] J. Bertoldo, E. Decencière, D. Ryckelynck, and H. Proudhon, *Glass fiber-reinforced polyamide 66 3D X-ray computed tomography segmentation segmentation U-Nets*, eng, Mar. 2021. DOI: 10.5281/zenodo.4601560.