

BENOIT BAZARD

SOFTWARE DEVELOPMENT ENGINEER

Table of contents

Professional experiences.....	2
Amadeus 2018 – 2020 (18 months).....	2
Internship Dassault Aviation 2017 (2 months).....	3
Thesis on deep learning 2016 – 2017 (10 months).....	4
Internship Outscale 2015 (3 months).....	6
School projects at Telecom SudParis.....	7
Chess engine with artificial intelligence 2015 (3 months).....	7
phase-contrast microscopy 2016 (3 months).....	8
Space Invaders 2016 (2 weeks).....	9
Personal Projects.....	10
Openid-connect 2019.....	10
Cryptopals Challenges 2018.....	10
Nextcloud and calendar synchronization 2018.....	11
Personal Website 2017.....	11
Random Notes Generator 2015.....	11
Education.....	12
Languages.....	12

Contact

Email: benoit.bazard@gmail.com
Personal website: bbazard.github.io

PROFESSIONAL EXPERIENCES

AMADEUS

2018 – 2020 (18 MONTHS)

Amadeus is a major Spanish IT provider for the travel industry.

Context:

In the team monitoring transactions concerning the booking flow of flights, working on huge data flows, the biggest exceeding 20 Terabytes per day. The MySQL database of the team contains 6000 tables. The team is composed of 5 developers.

I performed on my own the following activities:

- Technical migrations of the projects of the team after external dependencies changed:
 - From memcached to couchbase (needing authentication).
 - From a version of SLES (SUSE) to another one.
- Migration of a python project to use anaconda (from hard-coded python dependencies).
- Setup of a new data flow with more details for investigation purposes with hadoop, spark, hive.
- New flow added in an existing kafka topic, in C++.
- Sizing of a new kafka topic.
- Creation of a new REST service for a flow in python and flask.
- Work on the UI in Angular 5 to make it more consistent.
- Setup of a monitoring dashboard with non-regression tests in Jenkins pipeline.
- Incident investigations, some examples:
 - Filesystem full → Synchronization files of a data flow never deleted and accumulating.
 - Connection problem → Another team added authentication on one of our dependencies.
 - Connection problem → IP of a database that changed.
 - Numbers suddenly the same in all reports → Data type of MySQL field being too small.
 - Table empty in MySQL → MySQL partitions not created because of a software load forgotten.

There were also some projects involving the whole team:

- Study of how to migrate a MySQL database to ElasticSearch.
- Migration to Openshift/Kubernetes of an aggregation flow for transaction statistics.
- Onboarding for newcomers.
- Taking turns, being the point of contact of the team.

Technical environment:

C++, Python, Java

MySQL, Kafka, ElasticSearch, Couchbase

Openshift, Kubernetes,

Hadoop, Spark, Hive,

Angular

Dassault Aviation is a major player to aeronautics.

Context:

Using machine learning techniques for automated detection of failures in aircraft sensors.

- Presentation of machine learning and how it can be used for the department performing tests during flight (20 slides).
- Discussion with different parts of the organization to better understand how to recognize the sensor failures in the aircraft.
- Exploration of the existing techniques already used in the company to automatically detect the easiest sensor failures and decision of where (which sensor, which aircraft, which flight) machine learning techniques would be the most beneficial.
- Setup of an architecture of an automated sensor failures detection framework.
 - Choice of which machine learning algorithms would be most suited for the automated sensor failures detection.
 - Data warehouse creation– the different data flows were not centralized and in a data format created by the company.
 - Data manipulation to have data suited to the machine learning algorithms.

The implementation could not be done due to personal reasons.

Technical environment: C, Python

Context:

Identifying license plate character segmentation boundaries using convolutional neural networks, in python, for a Thai company specialized in car traffic monitoring.

This thesis is part of my dual degree at the Asian Institute of Technology, Thailand.

State of the art study:

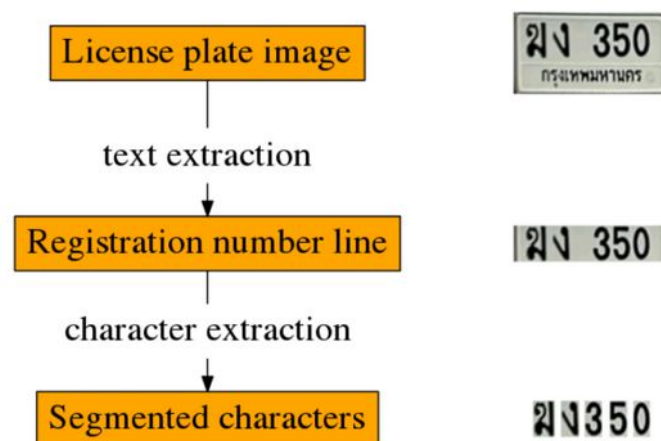
The following techniques have been studied:

- Region-Convolutional Neural Network (R-CNN)
- Pixel Projection
- Hidden Markov Chain (HMM)
- Component Analysis
- Long Short Term Memory (LSTM)

Methodology:

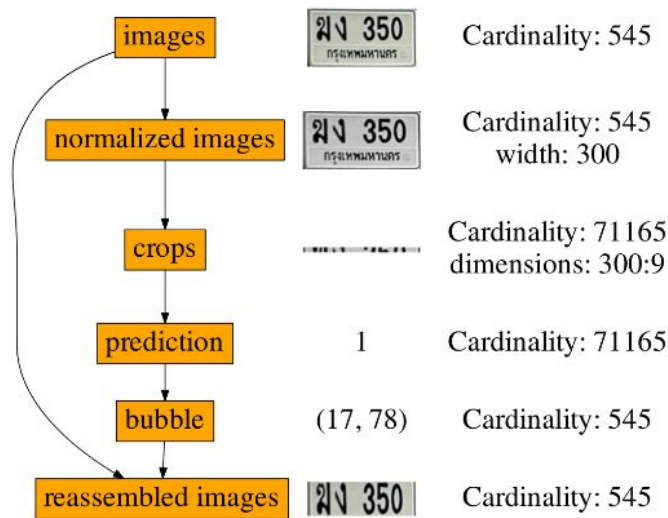
For a thesis, the approach must be innovative, thus it's not possible to only implement one of the already existing techniques commonly used. For that reason I had to come up with a new approach. To this end I took inspiration from the techniques of R-CNN and Pixel Projection.

First, there is an extraction of the main line of the license plate. In a second time, from that line, we extract each character separately.



The character extraction is divided in two phases.

For each extraction, the license plate is cut into thin overlapping slices. Each slice goes through a neural network that label them into good (1) or bad (0), the extraction is rebuild from the good slices.



*Here, the main line extracted is between the pixel rows 17 and 78 of the original image.
The slice prediction is good (1).*

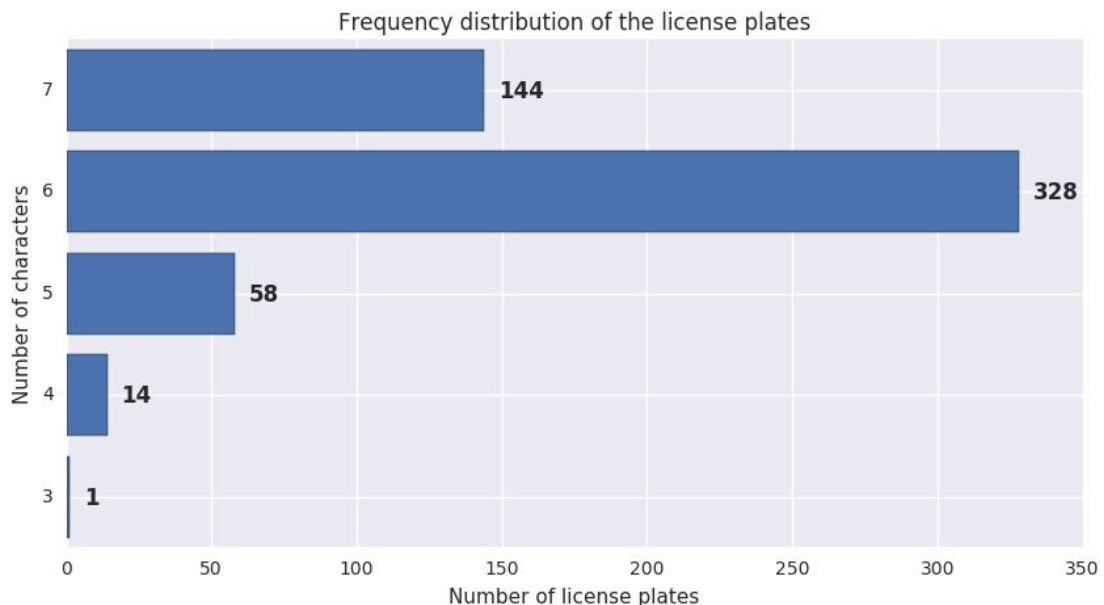
The 545 licenses plates generate 71 165 slices for the first phase (main line extraction).

A similar process is used to extract the individual characters from the main line.

For more information and pictures, see the completed thesis:

<https://github.com/BBazard/thesis/blob/master/bbazardThesis.pdf>

In order to better understand the license plate dataset provided to train the neural networks, I created several visualizations. This dataset was composed of 545 license plates.



The license plates have mostly 6 characters.



The layout of the characters in a license plate are not standardized.

On this last image, each line is a license plate; each colored segment is the room taken by a character horizontally. The colors are only there to better distinguish the different characters.

We can see that among the license plates with 6 characters, there is several clearly different layouts. The second line (frame31) has three characters on the left and three characters on the right whereas the last line (frame 452) has two characters on the left and four on the right.

The matplotlib source code for the visualizations is available at:

<https://bbazard.github.io/visualization>

Technical environment: Python (numpy, pandas, matplotlib, caffe, scikit-learn), Anaconda, NVIDIA DIGITS, Jupyter, LaTeX

INTERNSHIP OUTSCALE

2015 (3 MONTHS)

Outscale is a french Cloud Provider.

Context:

Discovery of Big Data, Cloud, and DevOps processes.

- Conception of a marketing document presenting Cloud and Big Data.
- Hadoop (MapR) installation on the cloud, distributed on three virtual machines via SSH.
- Deployment of virtual machines with Docker, Vagrant, VirtualBox.
- Deployment of Nginx in the cloud, and redaction of a how-to for customers.
- Utilization of boto (python interface for Amazon EC2, compatible Outscale).
- Website Scraping (Beautiful Soup, Selenium).
- Discovery of tools: Jira, Ansible, Jenkins.

Technical environment: Python, Beautiful Soup, Selenium, Docker

SCHOOL PROJECTS AT TELECOM SUDPARIS

CHESS ENGINE WITH ARTIFICIAL INTELLIGENCE

2015 (3 MONTHS)

Context:

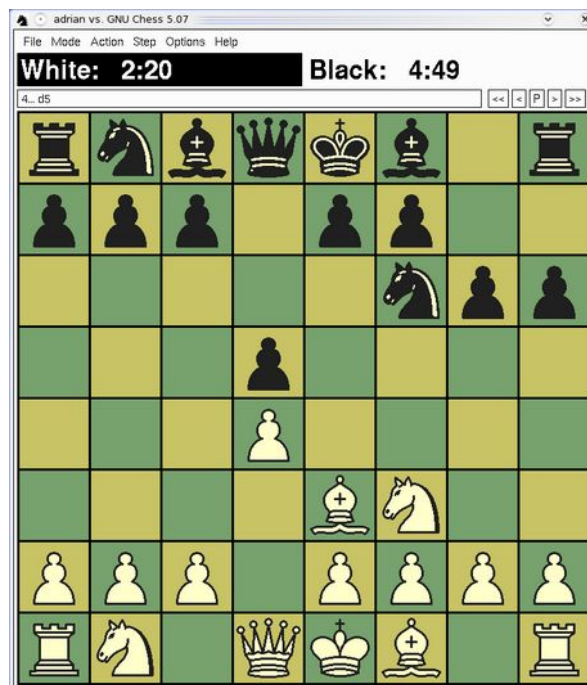
In a group of four, implementation of a min-max algorithm with some heuristics to determine how favorable is a certain chess configuration.

Utilization of the UCI protocol to be able to play with a graphical interface against other chess engines, thanks to the two following tools:

- XBoard - X Window System Chess Board
- Polyglot - Chess engine protocol adapter

Xboard can display a graphical interface which allows us to play against our artificial intelligence.

Polyglot is an Xboard adapter that allows us to send the next move and receive the opponent (our artificial intelligence) response.



Screenshot of the GUI XBoard.

In addition to developing the program, I was also the team leader. I distributed the workload taking into account everyone's skills and affinities. I enforced the usage of good practices such as git, valgrind, lint and was doing peer programming with the less experienced members of the team.

<https://github.com/BBazard/chessaint>

Technical environment: C, Subversion, Doxygen, Valgrind, gdb, cpplint.py

Context:

A phase-contrast microscope is able to take several images of a same object, each image is taken with a different phase. There is a way to combine all these images to create one very high resolution image. To that end, an algorithm converts the images to the Fourier Domain and then back with some intermediate steps.

In a group of two, implementation of this algorithm: the Gerchberg-Saxton algorithm.

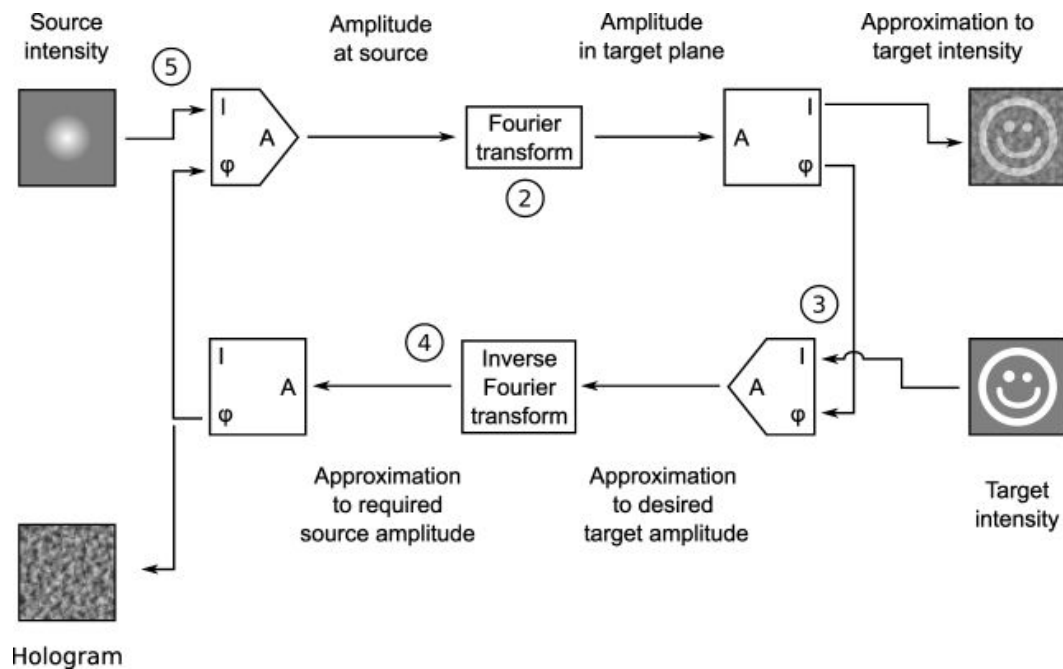


Illustration of the Gerchberg-Saxton algorithm.

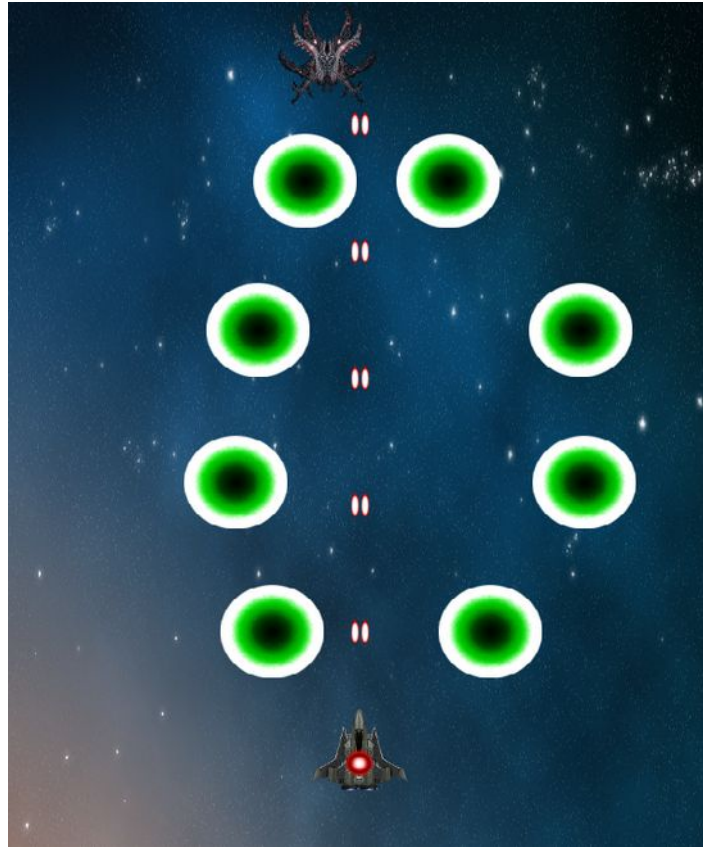
<https://github.com/BBazard/fourierscope>

Technical environment: C, openmp, libtiff (tiff format), libfftw3 (Fourier Transform), Google Test, Doxygen

Context:

For a course about the C++11 smart pointers.

In group of two, implementation of a space invaders clone with the design pattern Entity Component System. Some projectiles are following Lagrange curves to make them less predictable.



Screenshot of the game.

The sprite images come from internet.

<https://github.com/BBazard/INTouhou>

Technical environment: C++

PERSONAL PROJECTS

OPENID-CONNECT

2019

Context:

Protection of a REST api using keycloak, an Identity and Access Management (IAM) server.

- Access token and ID token handling.
- Understanding of the different grant types: Implicit, Authorization code, Hybrid, Client credentials, Resource owner password, Refresh tokens.

Technical environment: Python (Flask), Keycloak, docker-compose

CRYPTOPALS CHALLENGES

2018

Context:

Set 1, 2, 3, 4 of cryptopals.com, cryptographic challenges.

Implementation of several cryptographic primitives and attacks including:

- Detection oracle
- Padding oracle
- Bitflipping attacks
- Byte-at-a-time ECB decryption
- Crack an MT19937 seed
- Clone an MT19937 RNG from its output
- Break a SHA-1 keyed MAC using length extension

Crypto Challenge Set 3

This is the next set of **block cipher cryptography** challenges (even the randomness stuff here plays into block cipher crypto).

This set is **moderately difficult**. It includes a famous attack against CBC mode, and a "cloning" attack on a popular RNG that can be annoying to get right.

We've also reached a point in the crypto challenges where all the challenges, with one possible exception, are valuable in breaking real-world crypto.

17. The CBC padding oracle
18. Implement CTR, the stream cipher mode
19. Break fixed-nonce CTR mode using substitutions
20. Break fixed-nonce CTR statistically
21. Implement the MT19937 Mersenne Twister RNG
22. Crack an MT19937 seed
23. Clone an MT19937 RNG from its output
24. Create the MT19937 stream cipher and break it

Screenshot of the set 3.

https://github.com/BBazard/cryptopals_challenges

Technical environment: Python

Context:

Setup of a nextcloud instance on a private local network (ip like 192.168.*.*) with setup of a local trusted certificate authority to have a calendar synchronization with SSL between Linux computer and Android smartphone.

Technical environment: Docker, Apache, OpenSSL

PERSONAL WEBSITE

2017

Context:

Creation of a website portfolio: bbazard.github.io

Technical environment: HTML, CSS, JavaScript

RANDOM NOTES GENERATOR

2015

Context:

LilyPond snippet that randomly generates notes on a music sheet to train sight reading.



Line of a pdf generated by the script.

<https://github.com/BBazard/random-notes-generator>

Technical environment: Scheme (Lisp), Lilypond

EDUCATION

2014 – 2018	Engineer Diploma of Télécom SudParis Évry, France
2016 – 2017	Master of Science of the Asian Institute of Technology Bangkok, Thailand
2011 – 2014	Classes préparatoires aux grandes écoles - PT Lycée Chevrollier – Angers, France
2011	Baccalauréat scientifique, specialty mathematics Lycée David d'Angers – Angers, France

LANGUAGES

FRENCH	Native Proficiency
ENGLISH	Full Professional Proficiency (TOEIC 955/990 in 2016, TOEFL 97/120 in 2015)
JAPANESE	Intermediate Proficiency