IMAGE PROCESSING APPLIED TO DETECTION OF MULTIPLE OBJECTS IN PUBLIC TRANSPORTATION  
  
**Place of Final Internship**   
Orange Applications for Business   
**Tutor of Final Internship**Lionel FREMANTEAU  
**INSA Educational Supervisor**Jean-François NEZAN

**PFE presented on 08/28/2018**

Final Internship  
presented by  
**Marie BRUNET CARTEAUX**  
Engineering student from INSA Rennes  
Speciality EII  
Scholar year 2017-2018

**PFE**

**Table of contents**

[Acronyms 3](#_Toc516501694)

[Acknowledgments 4](#_Toc516501695)

[Introduction 5](#_Toc516501696)

[I. Context 6](#_Toc516501697)

[I.1. Company 6](#_Toc516501698)

[I.1.1. Orange Business Services 6](#_Toc516501699)

[I.1.2. Orange Applications for Business 6](#_Toc516501700)

[I.1.3. Machine-to-Machine Department 6](#_Toc516501701)

[I.2. Problematic 7](#_Toc516501702)

[I.2.1. Bid solicitation 7](#_Toc516501703)

[I.2.2. Constraints 7](#_Toc516501704)

[I.2.3. Existing work 8](#_Toc516501705)

[I.3. Objective 8](#_Toc516501706)

[II. Solution 8](#_Toc516501707)

[II.1. State-of-the-art 8](#_Toc516501708)

[II1.1. Semantic segmentation 8](#_Toc516501709)

Table of figures

[Figure 1: Example of bad light exposure. An underexposed image presents low contrast and produces a histogram with mainly low frequencies, while an overexposed image is mainly composed of high frequencies. 7](#_Toc518288013)

# Acronyms

|  |  |
| --- | --- |
| **B-B** | Bounding Box(es) |
| **BGR** | Blue Green Red |
| **CNN** | Convolutional Neural Network |
| **DBN** | Dynamic Bayesian Model |
| **DL** | Deep Learning |
| **DNN** | Deep Neural Network |
| **FSF** | Functional Specifications File |
| **GDPR** | General Data Protection Regulation |
| **HSL** | Hue Saturation Light |
| **IoT** | Internet of Things |
| **MTM** | Machine To Machine |
| **OAB** | Orange Applications for Business |
| **OBS** | Orange Business Services |
| **POC** | Proof of Concept |
| **SaaS** | Software as a Service |
|  |  |

# Acknowledgments

I would like to thank the following people for the great experience they offered me during this internship: Thierry Drouet for offering me the opportunity of doing an internship at Orange Application for Business, my tutor Lionel Frémanteau, for he always made time for me when I needed help or struggled in making a decision. I discovered great integration tools and got invaluable advice from him about best practice in C++. Regards to my manager Florence Laup, for she valued my job even when I lacked of self-confidence.

I would like to greet all my colleagues, for they made this internship a very pleasant time. It was a pleasure to work in their company.

My thanks go to my educational supervisor Jean-François Nezan, for his support and responsiveness.

Finally, I would like to dedicate this report to my late foster father, who was proud I would become an engineer.

# Introduction

Though it has already proved its efficiency and usefulness in the last decades, machine learning is still a innovating technology and continues to revolutionize many industrial domains. Supported by the growing performance of hardware devices and the cheapness of cloud services, deep learning emerged in the mid-2000s as a complex technology based on numerous hidden computational layers, able of detecting features in images and videos amongst hundreds of categories.

In a context of sustainable development, Orange Business Services (OBS) aims at developing solutions for detecting people and abandoned objects in public transportation, for it could improve the transportations companies throughput and safety onboard. Such problematic is of great interest for numerous companies, thus OBS intends to be prepared to respond to similar tenders.

Three main constraints compel real-time multi-object detection, namely the hardware requirements, the varying quality in images and the sensitivity of travelers-related data. This report exposes the undertaken research and development to address the issue of multi-object detection with regard to those constraints.

The context of the internship is firstly presented. A state-of-art of the existing techniques (either related or not to deep learning) for detecting people and objects in images is drawn in the second part, before disclaiming the technological choices that have been made for developing the solution. The development pipeline will then be extensively approached in part three alongside with adopted management and design strategies. Results and potential improvement will finally be discussed in part four.

# I. Context

## I.1. Company

### I.1.1. Orange Business Services

Founded in 2006 as the business services arm of Orange S.A., Orange Business Services is a global integrator of communications products and services for multinational corporations. Operating in 220 countries and territories with the help of over 21 000 employees, OBS has more than 8000 registered patents in their portfolio and produced a revenue of €6.4 billion in 2015. The company specializes in Internet of Things (IoT), smart cities, digital transformation, big data analytics and artificial intelligence, and invested 700 million euros in research and development over the last few years.

### I.1.2. Orange Applications for Business

Specialized in systems integration and providing custom-made or SaaS applicative services, Orange Applications for Business support companies throughout their projects lifetime (counseling, conception, development and operation) in three main fields:

* **Client experience:** with solutions for all steps of customer path before, during and after purchases, on all communications channels (digital, vocal, local)
* **Big data/analytics:** to make available critical data that help companies in making strategical and operational decisions
* **Machine-to-machine and smart objects:** to help companies make their objects communicate so as to improve efficiency and support development of their activity.

OAB is present on three sites in Rennes area and employs over 2300 people around the world. It offers services for more than 20 000 companies and generates €465 million per year.

### I.1.3. Machine-to-Machine Department

MTM is the department for machine-to-machine, real-time and multimedia projects. It integrates telecommunications and informatics technologies and deploys solutions for communication between machines and people without the need of human intervention.

MTM department proposes interconnection solutions to increase the customers efficiency in their business processes and creates new use-cases in industry, logistics and trade with the help of seamless smart objects.

## I.2. Problematic

### I.2.1. Bid solicitation

Detection of features in images has become something companies are eager to set up. Many tasks currently performed by humans are deconstructed and carefully studied in order to understand what makes it so complex to automatize them. Hopefully computers are increasingly becoming faster and more performant, enabling the release of algorithms and smart networks that can recognize multiple objects in images and videos from up to a thousand categories of objects.

Orange Applications for Business plans to develop a solution for identifying features in images from different use-cases, *e.g.* counting humans in public transportation, or identifying free lots, whether it is in warehouse shelves or in a parking. By developing a solution answering this problematic, OAB could offer additional features to its existing solutions that would deeply improve the customer’s experience.

### I.2.2. Constraints

For the sake of simplicity, I jointly decided with my tutor to focus on the use-case of public transportation. This implies th at two main categories of objects have to be detected **and** recognized : people and abandoned objects (whether it is luggage, clothes, various bags …). Although it could look simple at first sight to classify features in three categories (*i.e.* 1: people, 2: abandoned objects and 3: everything else), many constraints add up to the classical image datasets on which these algorithms and networks train. Such constraints can be:

- **Changes in the light exposure (cf Figure xx), contrast, or position** of the camera. The designed solution shall be robust and generic enough to handle all those changes.

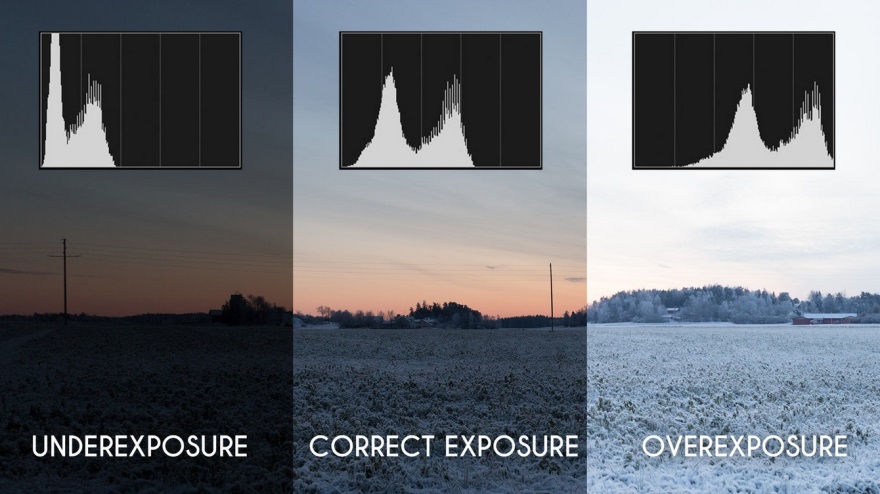


Figure 1: Example of bad light exposure. An underexposed image presents low contrast and produces a histogram with mainly low frequencies, while an overexposed image is mainly composed of high frequencies.

- The **period** at which the solution shall be executed. As this will depend on the use case, the goal is to conceive a solution **as fast as possible**.

- The mandated **execution latency** of the solution. This latter shall not run continuously because of the useless data and resources consumption.

- The **initialization needs**. Since most use-cases aim at rolling out the solution on a substantial number of devices, it is crucial to shrink the initialization needs as much as possible.

- **Reuse of video equipment** already deployed for other services and/or hardware resources of cameras, that do not come with GPUs. This prevents the processing of data on the spot and implies the usage of **remote hardware resources** (cloud computing).

- **Data anonymization** asked by the European referential for customer privacy (otherwise known as the GDPR, who came into effect last May). The solution shall either ensure the **complete anonymization** of data, such that it is impossible to identify anyone from the data, or not **store** any sensitive information.

### I.2.3. Existing work

There is no solution developed by OAB addressing the exact problematic as abovementioned to this day. Collaborators from another Orange business unit devised a method for detecting people in buses based on template matching, but it requires human intervention to get prior knowledge of the background. This is not a conceivable industrial solution since the “scan” of the vehicle and replica making consumes too much time and computing resources. On one hand, this project is currently being upgraded to get rid of those initialization needs, which will make the solution way faster and lighter. On the other hand, Orange wishes for an algorithm able of detecting both people and abandoned object at the same time.

The main objective of this internship is thence to perform a proof of concept (POC) for automatically detecting and recognizing people and abandoned objects in different vehicles of public transportation. This solution shall be robust to light, exposure and camera position changes and must run periodically.

# II. Solution

## II.1. State-of-the-art

The first step of this internship was to perform a state-of-the-art mapping out the existing techniques for detection of multiple objects in frames. This document would help the further writing of the design brief and associated functional specification documents.

Though image processing is the generic term for any method or algorithm aiming at using or modifying images and therefore theoretically includes deep learning, these two notions have been considered differently in this document for the sake of simplicity. Consequently, image processing refers to any tool that does not involve artificial intelligence.

There is no mention in the literature of algorithms addressing this exact issue. It was pointed out, however, that two main categories of algorithms were particularly promising to perform this task, namely semantic segmentation image processing-based techniques, and deep learning-based techniques.

### II.1.1. Semantic segmentation VS object detection

Image processing- and deep learning-based methods do not rely on the same techniques and postulates.

Segmentation is a subset of images processing that aims at extracting features of interest in images. Several types of segmentations exist, which serve different purposes:

* **Semantic segmentation** looks for consistency in adjacent objects. On Figure xx below, the segmentation algorithm recognizes the **continuation** of the chairs even if the table cuts the image in two parts. However, since the chairs are close enough to each other, the algorithm does not recognize the objects’ uniqueness and consider them as a whole.
* In **boundary segmentation**, the algorithm looks for **uniform areas** in the image and is not able to **grasp the continuation** across zones of discontinuity (for instance, in Figure xx, the chairs in the left of the image are segmented as two different parts around the table).
* In **semantic instance segmentation**, principles of semantic segmentation are kept, and the idea of **uniqueness** is added. The algorithm is able to detect the very little distance between the objects and segment them individually (see on the right image in Figure xx).

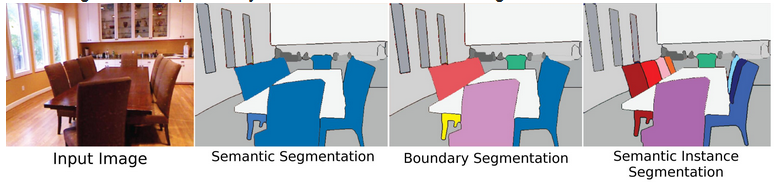
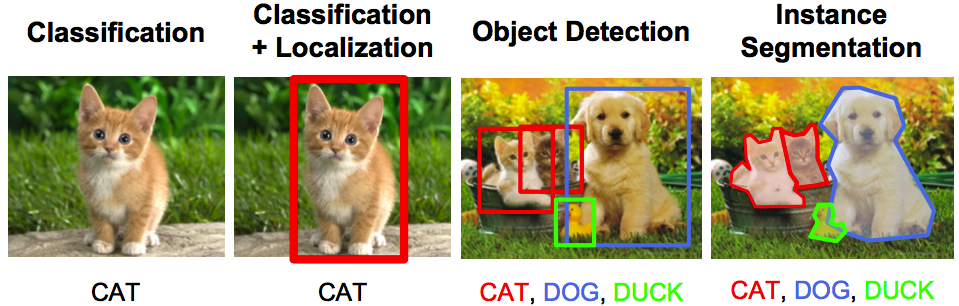


Figure 2: Different types of segmentation.

In the case of these studies, the most interesting type of segmentation to set up would be semantic instance, since it enables distinguishing individuals and have a precise idea of their number.

Similarly, there is different types of object detection, as described below:

* Classification analyses the image content and gives



**Détection automatique d’usagers et d’objets abandonnés dans les transports en commun**

**Autonomous multi-object detection in public transportation**

Marie BRUNET CARTEAUX

**Résumé**

Si le machine learning avait ouvert la voie dans les années 1960, son descendant l’apprentissage profond est en train de révolutionner la résolution autonome de problèmes impliquant des objets aussi complexes que des images ou des vidéos. Fascinant à la fois les acteurs de la recherche et de l’industrie, l’apprentissage profond promet d’adresser de nombreuses problématiques, allant de la classification de cellules cancéreuses à l’interprétation de la signalétique routière.

Ce rapport de projet de fin d’études propose une solution de détection automatique d’usagers et d’objets abandonnés dans les transports publics. Cette solution vise à terme à permettre le renforcement de la sécurité à bord des véhicules de transport en commun et d’optimiser le placement des véhicules sur l’ensemble du réseau.

Une étude de l’état de l’art a été faite afin d’analyser les solutions de détection multi objets déjà existantes. La solution proposée a été développée en C++ avec l’utilisation de la librairie de traitement d’image OpenCV et la bibliothèque de réseaux profonds de neurones TensorFlow.

**Mots-clés :** Apprentissage profond, Réseaux de neurones, Traitement d’image, Optimisation, Développement durable, Sécurité.

**Abstract**

Though machine learning has already proved its usefulness and efficiency in the last decades, Deep Learning, which places itself as its heir, is currently revolutionizing autonomous learning and optimization. Able of solving problems involving complex and heavy objects like images and videos, deep learning addresses an extensive range of problematics, from the cancerous cells identification to road signs interpretation.

This reports aims at offering a solution for automatically detect people and abandoned objects onboard of public transportation vehicles. Such a solution could help enhancing safety onboard and optimize placement of vehicles on the transportation network.

Related work has been studied in a state-of-art to explore potential loads on multi-object detection. The final implementation is done in C++ and uses the image processing library OpenCV and deep learning framework TensorFlow.

**Keywords:** Deep Learning, Neural Networks, Image Processing, Optimization, Sustainable Development, Safety.