



# Multimodal learning for customs fraud detection & action recognition

---

Estimated duration : 4,30  
min

1 of 12

# Why we need Multimodal learning?

---



# Our experience of the world is multimodal

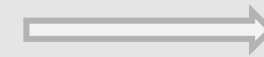
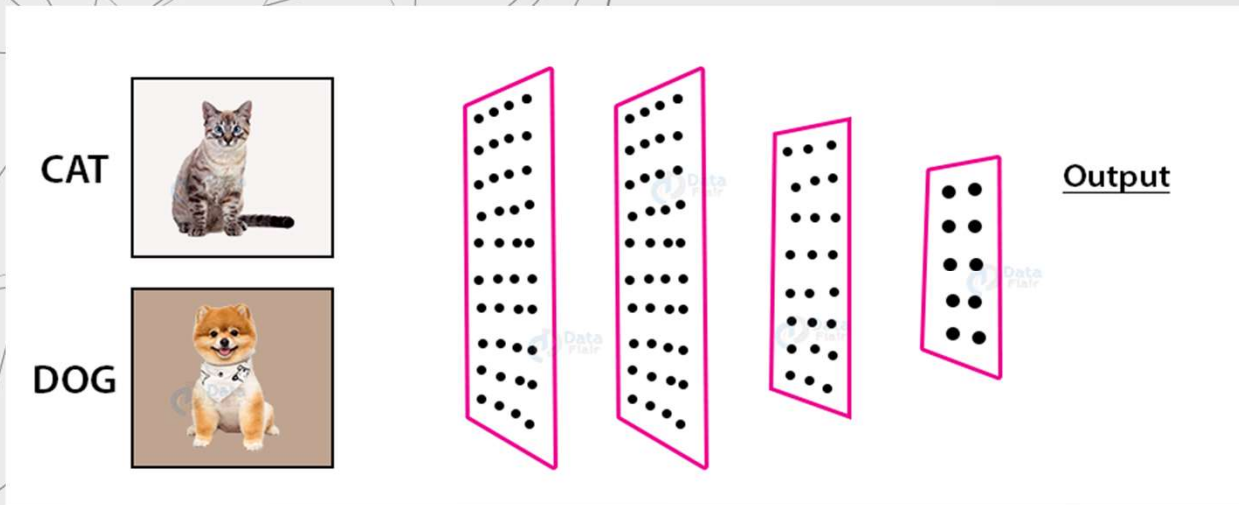
Multimodal learning suggests that when a number of our senses – visual, auditory, kinesthetic – are being engaged in the processing of information, we understand and remember more.



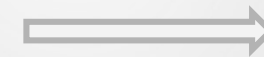
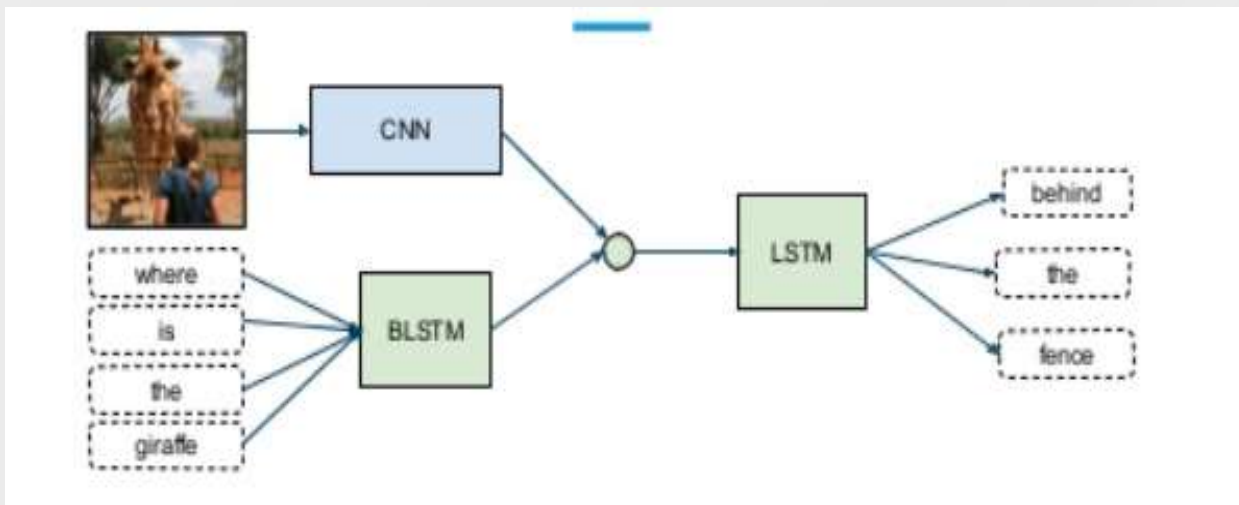
## Then what is a modality and multimodal learning ?

**Modality** : It is a way or an environment in which something happened or is experienced , **Examples** : image, audio , text...etc.

**Multimodal Learning** : When multiple modalities are involved during training and inference phase , we call that a multimodal learning



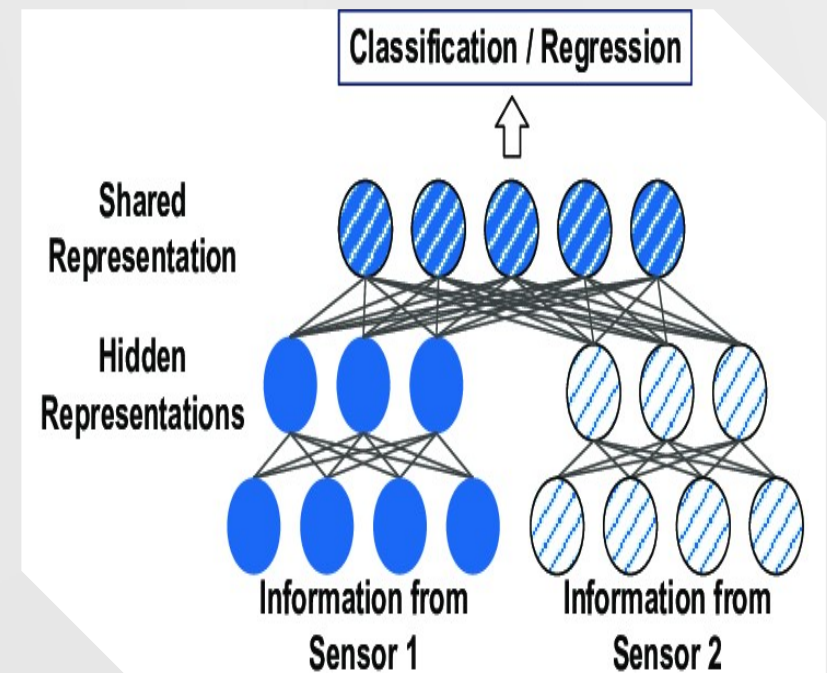
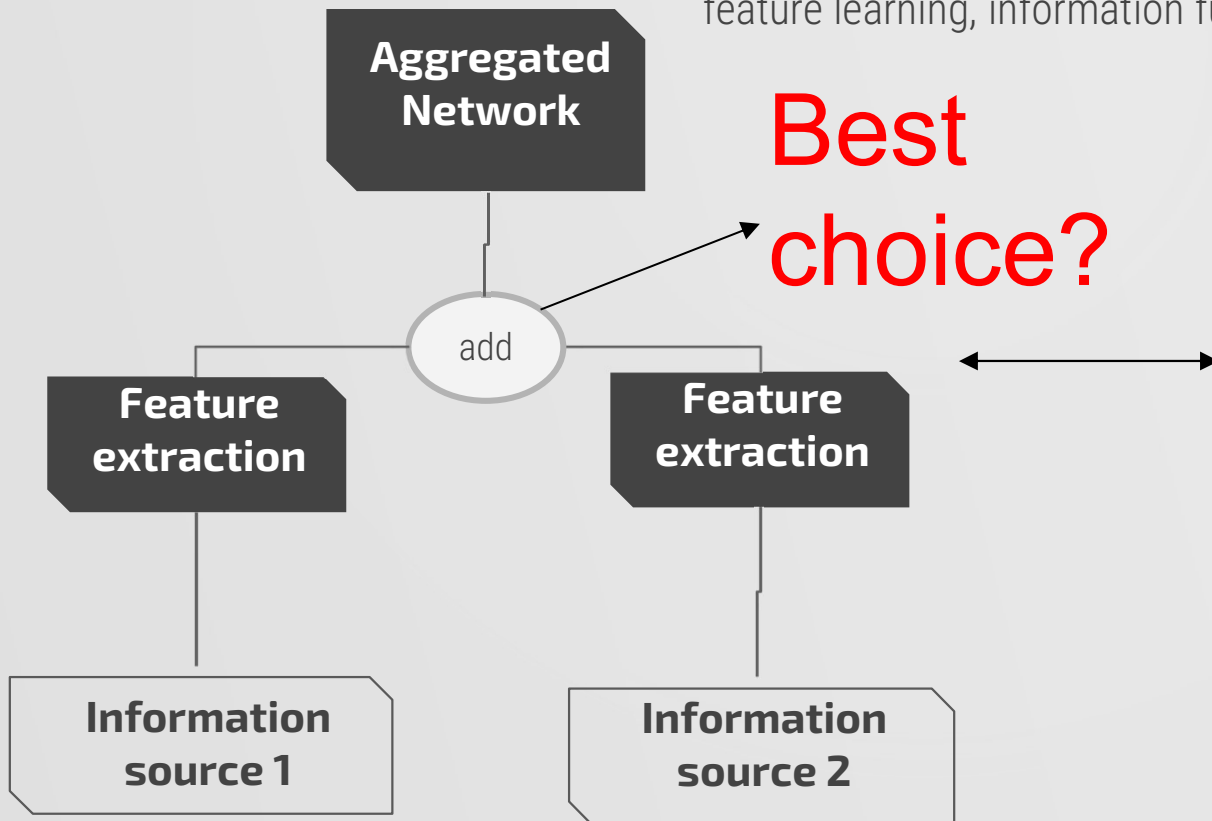
**Unimodal learning**



**Multimodal learning**

# How do they work

In multimodal learning, as its name suggests, we aim to do information fusion from different modalities to improve our network's predictive ability. The overall task can mainly be divided into three phases – individual feature learning, information fusion and testing.





# Main thesis objectives

01

Determine the appropriate **algorithm that is able to cope with multimodal learning** in context of multimedia processing and medical imaging

02

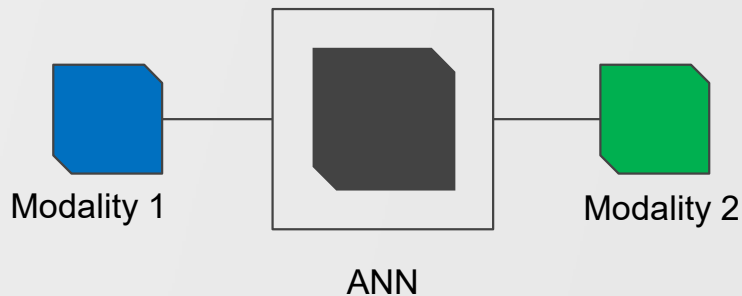
Exploit the **provided multimodal datasets from our sponsors (E-origin and Infrabel )** to validate and solve real-world problems

03

Build a framework capable of explaining our model results

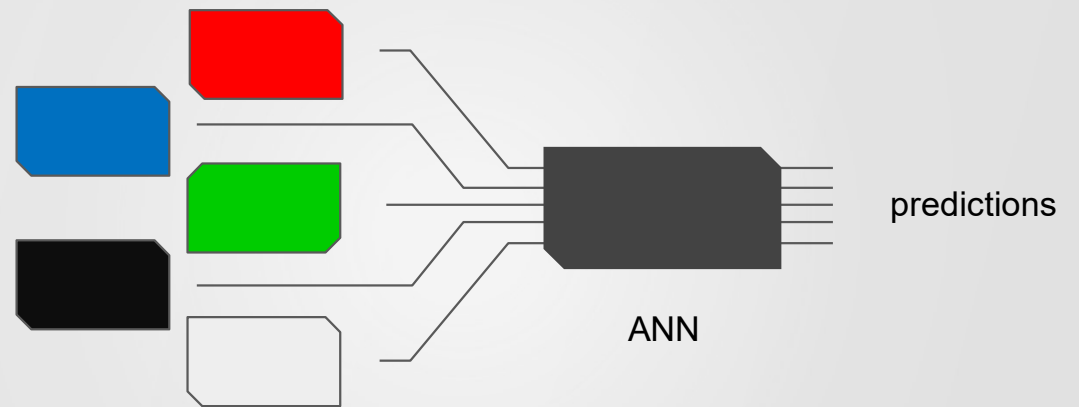
# Multimodal Learning can be divided into two categories :

## Modality to modality transition



Corresponding challenges are :  
Transition and Alignment

## Multi-modalities input



Corresponding challenges are :  
Representation , Fusion





# e-Origin

## Use Cases

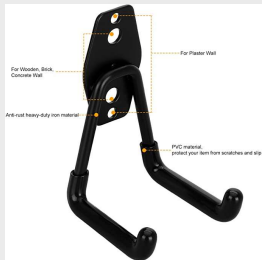
# INFR/ABEL

### Customs fraud detection

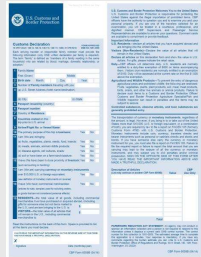


### Modalities

#### Images



#### Text (Customs déclarations , tarbel )



### Construction Site Worker's safety using AI



### Modalities

#### RGB Images



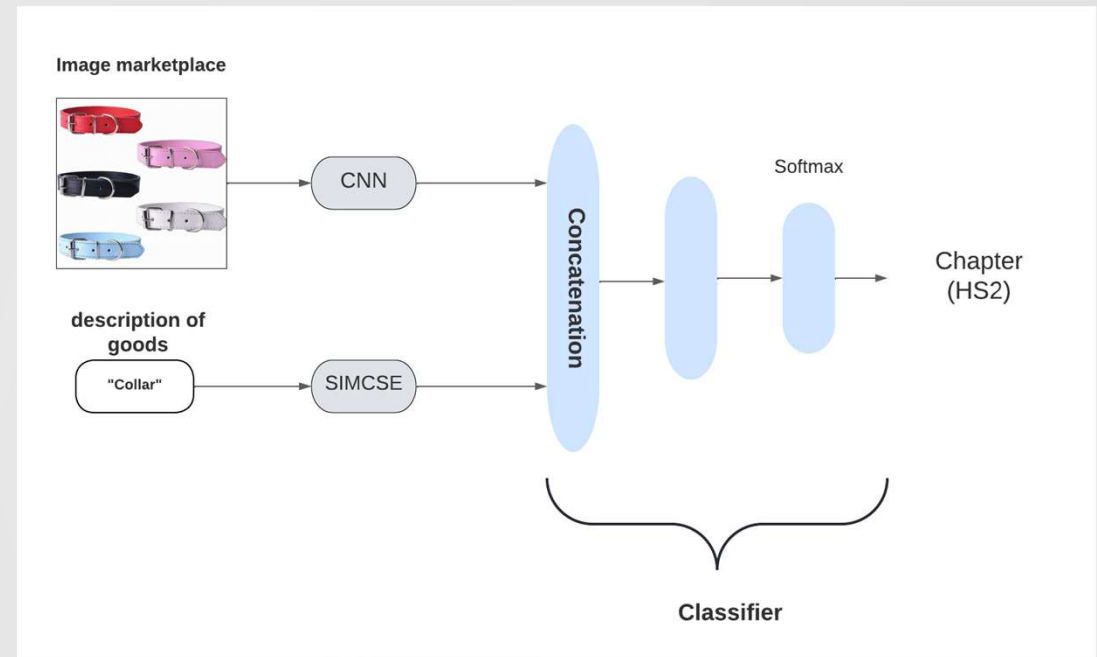
#### Depth maps



# First results

## Customs hs code prediction

Learning mode	accuracy
Text only (unimodal)	77.47 %
Image-text (multimodal)	83,51 %
Image-only (unimodal)	73,62 %





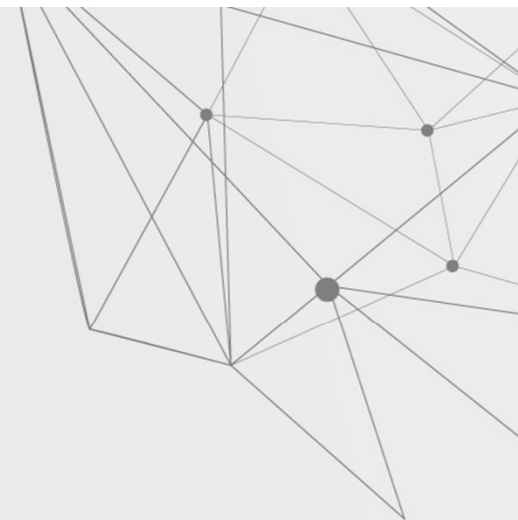
**UMONS**

**INFRABEL**



**e-Origin**

**numediart**



**Otmane AMEL**

Phd Student and Research Assistant

📍 Polytechnic Faculty – UMons

[Otmane.amel@umons.ac.be](mailto:Otmane.amel@umons.ac.be)

Supervised by :

**Sidi Ahmed Mahmoudi**



Co supervisor :

**Xavier Siebert**

