High evolution of heterogeneous data modalities during the last years

Unimodal Deep Learning solutions lacks an understanding of its surrounding environment

Determine the appropriate multimodal learning able to cope with multi-source of information

**Objective**

- Exploiting diverse data sources for action recognition and custom fraud detection

**Multimodal Taxonomy**

- **Fusion**
  - Late, early, Hybrid are model agnostic fusion

- **Co-Learning**
  - Learning enhancement using other modalities

- **Representation**
  - Representation: joint & coordinated

**Use Cases**

1- Customs fraud detection based on Hs Prediction

- **Dataset**: 1800 customs declarations supplied by *e-Origin*
  - Each declaration contains valid market-place URL
- **Image descriptor**: pre-trained VGG 16
- **Text descriptor**: pre-trained SIMCSE (bert-based)
- **Multimodal learning type**: Fusion, representation learning

**Results**: Hs Code Prediction

<table>
<thead>
<tr>
<th>Modality Type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text only Unimodal</td>
<td>77.47</td>
</tr>
<tr>
<td>Image-text Multimodal</td>
<td>83.51</td>
</tr>
<tr>
<td>Image only Unimodal</td>
<td>73.62</td>
</tr>
</tbody>
</table>

**Model architecture**

**Perspectives**

- Develop models able to cope with missing modalities
- Find the best combination between modalities to improve the model's performance
- Design explainability techniques for the developed models