TELECOM SudParis

INSTITUT POLYTECHNIQUE DE PARIS

# Utilising unlabelled data to enhance prediction, study case of gesture prediction.

HEDROUG Amine

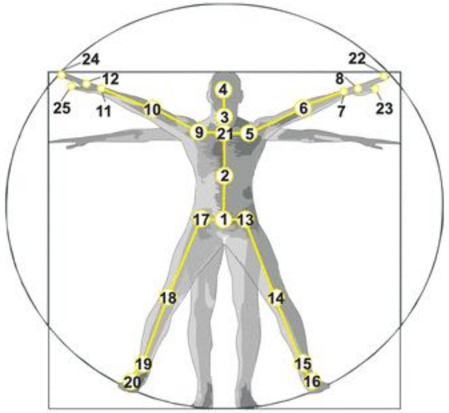
Institut Mines-Télécom

# Objectives

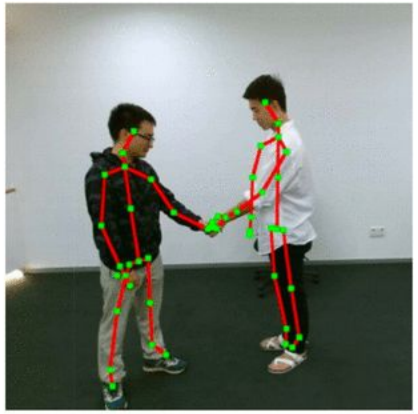
Objectives

# Objectives

- To try to use other learning mechanisms to use unlabelled data to enhance gesture based analysis:
  - Multiple Instance Learning (MIL)
  - Self-Supervised Learning
  - Semi-Supervised Learning



A


B

- The targeted task is gesture classification, but can be extended to other task, such as gesture generations etc.

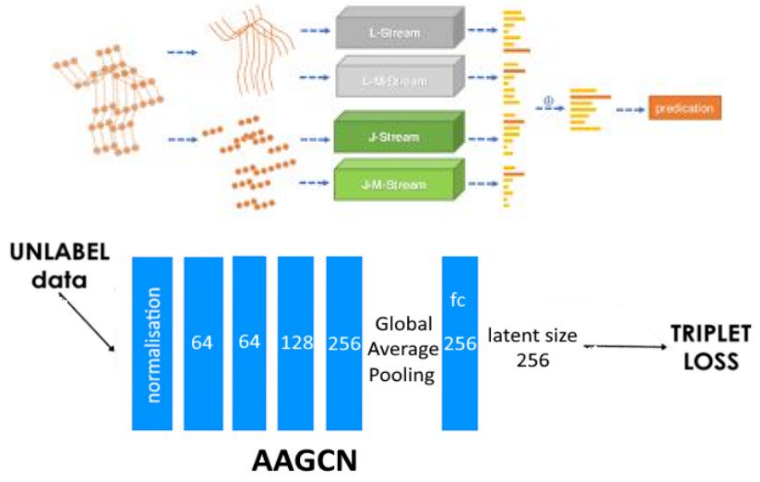
# Methods

Methods

# Methods

## Baseline methodes (CNN - Graph)

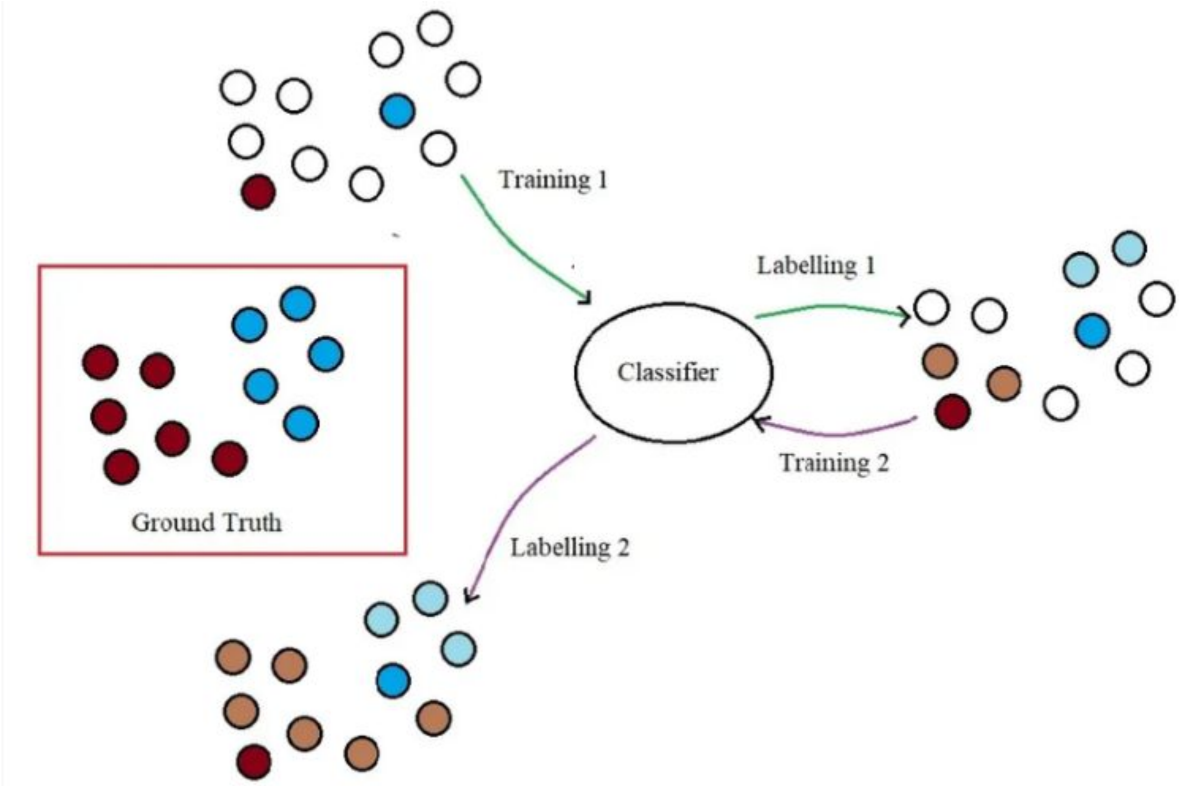
- Two baseline methods are conceived:
  - **CNN-based:** takes a grayscale image (rasterised table of joint sequences) as input, processes it with convolutions, and outputs class scores.
  - **Graph-based:** takes joint coordinates as a spatio-temporal graph, processes them with graph cnn, and outputs class scores via fully connected layers.
- The CNN based method is simpler, but can't consider the joint information.




# Methods

## Semi-Supervised

- **Train** a model on **labeled** data, then use it to predict labels for unlabeled data.
  High confidence predictions are added to the training set, and the process repeats iteratively.

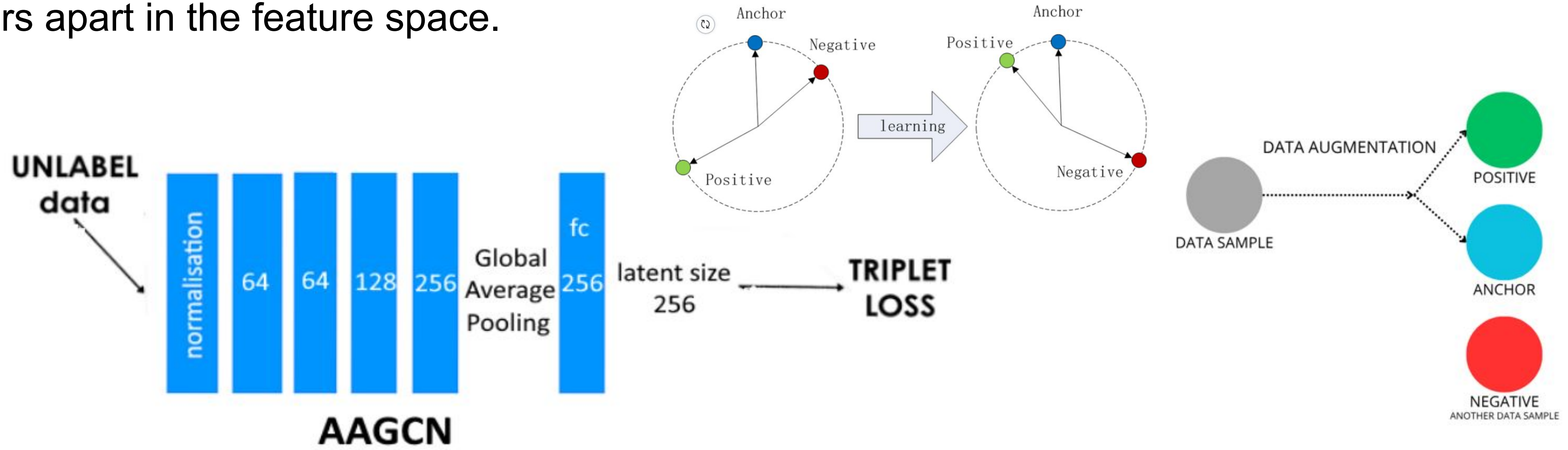



| DATASET BENCHMARK XSUB | |
|---|---|
| Train part : 40 000 | Eval part : 16 000 |
| 1% - 5% - 10% labeled / Unlabelled or Weakly Labelled | |

# Methods

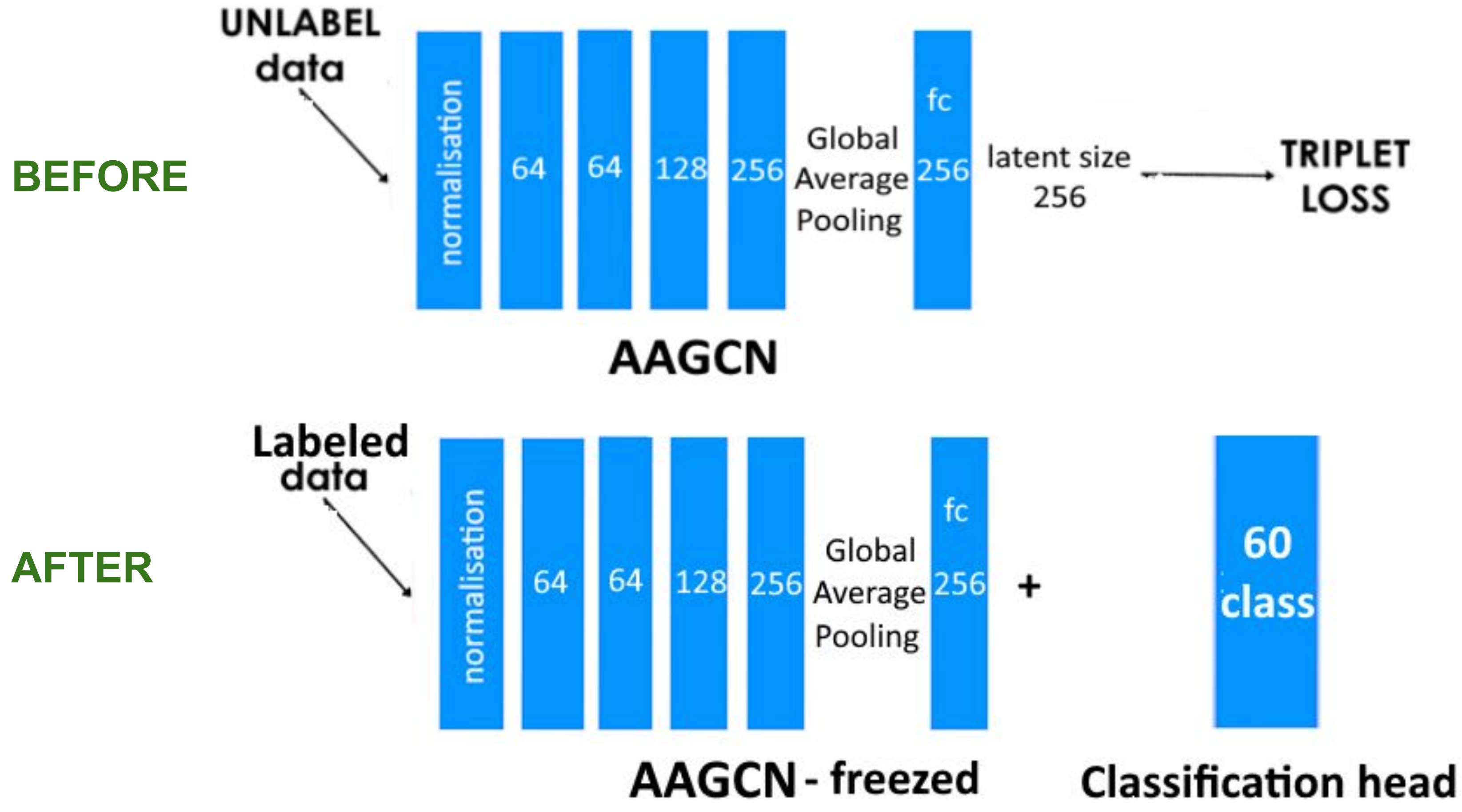
## Self-Supervised - Upstream Model

**Siamese contrastive learning :**
Contrastive learning is a self-supervised approach where a model learns to distinguish between similar and dissimilar data by bringing positive pairs closer and pushing negative pairs apart in the feature space.



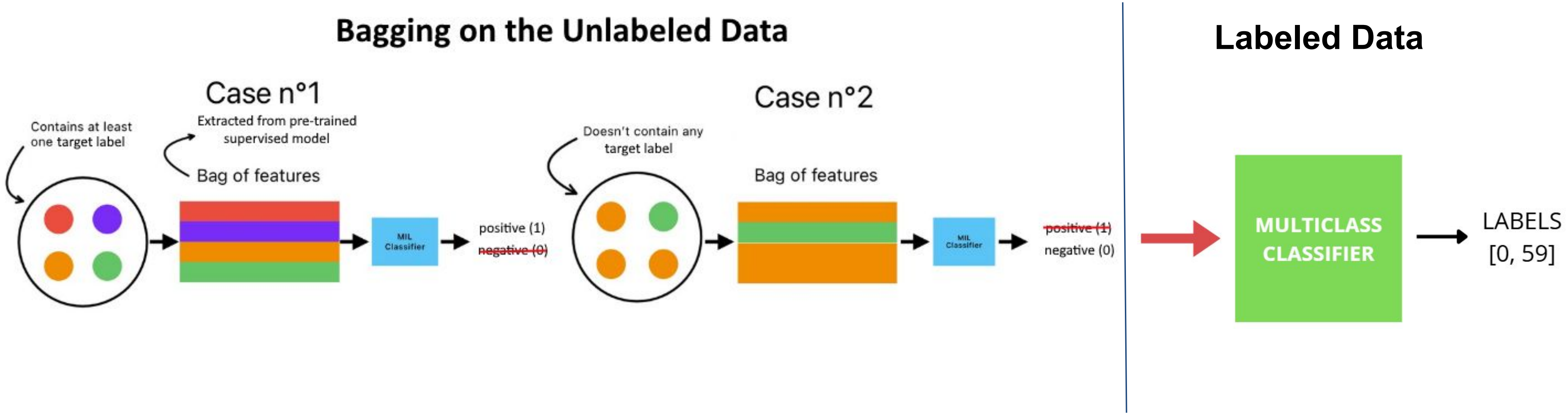
# Methods

## Self-Supervised - Upstream Model



# Methods

## Multiple instance learning




**Positif** : Daily actions and Medical conditions

**Negative** : two person interaction

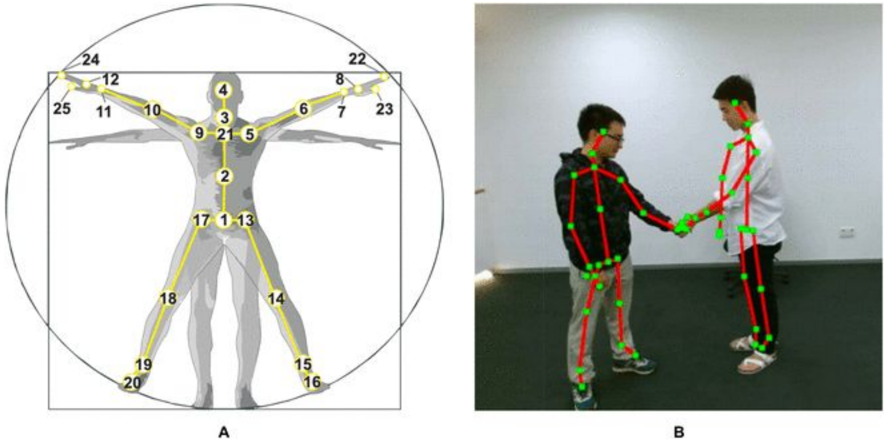
# Current State

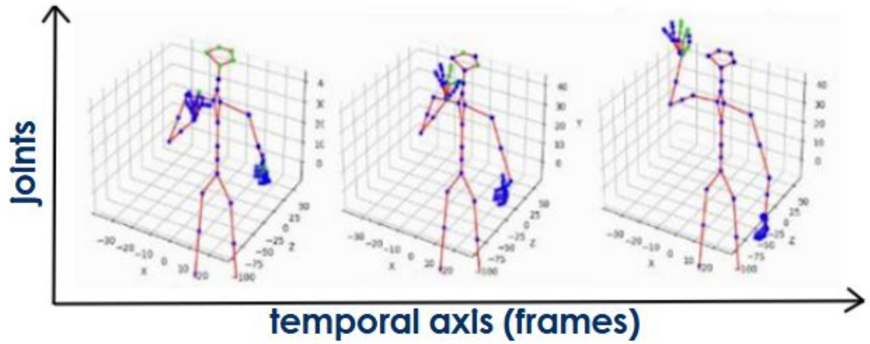
Current State

# Datasets

- **NTU RGB+D 60 dataset**:
  - Joints: 25x2
  - Class: 60 different actions
  - 56000 samples

- **Mocaplab dataset**:
  - Joints: 37
  - Class: face, torso, hands, other
  - 718 samples



# Current State

## Dataset

- Data splitting :
  - NTU RGB+D dataset (X-sub benchmark), focusing on low-label regimes with 1%, 5%, and 10% of the training set labeled.
  - Mocaplab: internally split the data into 50% labelled, 45% unlabelled, and 5% as test. All data split is fixed.

**DATASET BENCHMARK XSUB**

Train part : 40 000 | Eval part : 16 000

1% - 5% - 10% labeled | Unlabelled or Weakly Labelled

**718 Data**

Trainning Data | Validation

350 Labelled | 318 Unlabelled or Weakly Labelled | 50

Institut Mines-Télécom

# Current State

## Results

- Results on the mocaplab shows the improvements of standard supervised method upon introductions of our three methods, exploiting unlabeled data.
- For NTU dataset, the effect for semi-supervised method is seen, but for other two, are still in progressed.

### NTU RGBD

| | 1% | | 5% | | 10% | |
|---|---|---|---|---|---|---|
| | TRAIN (train+val) | TEST | TRAIN | TEST | TRAIN | TEST |
| **Supervised** | 8,59% | 7,61% | 38,48% | 34,31% | 55,31% | 51,77% |
| **Semi-supervised** | 10,53% | 9,49% | 43,86% | 40,09% | 63,03% | 57,91% |
| percentage of unlabeled data used | 5.63% | | 18.01% | | 55,55% | |
| **MIL** | in progress | | in progress | | 2,19% | 1,70% |
| **Self-Supervised** | in progress | | in progress | | 4,84% | 4,66% |

### Mocaplab

| Model | Accuracy | Relabelling accuracy |
|---|---|---|
| Supervised | 56% | - |
| Semi-supervised | 62% | 69% |

| Model | Accuracy | Loss |
|---|---|---|
| Supervised | 56% | 3.9 |
| Supervised + Contrastive (labelled) | 58% | 2.9 |
| Constrastive (unlabeled) + Supervised | 58% | 2.7 |

| Model | Accuracy | f1-score |
|---|---|---|
| Supervised | 68% | 0.69 |
| weakly-supervised | 70% | 0.63 |

# Observations

Observations

# Observations

## Semi-Supervised - NTU

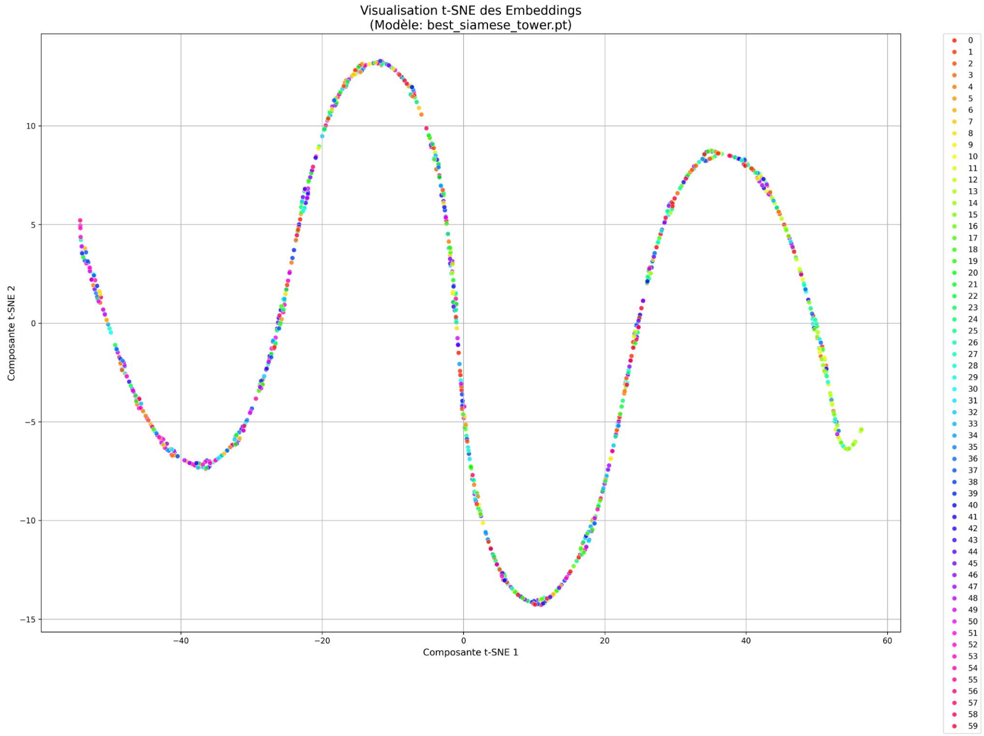
- We see that the semi-supervised learning allows to use the unlabeled data, that it introduces pseudo-classified samples as new information as training progress (for each phase).
- Here, 29.24% unlabeled data is now used as additional training samples, with 85.23% confidence level (i.e. out of 29.24% new samples, 85.23% are properly labeled - we have the true label)

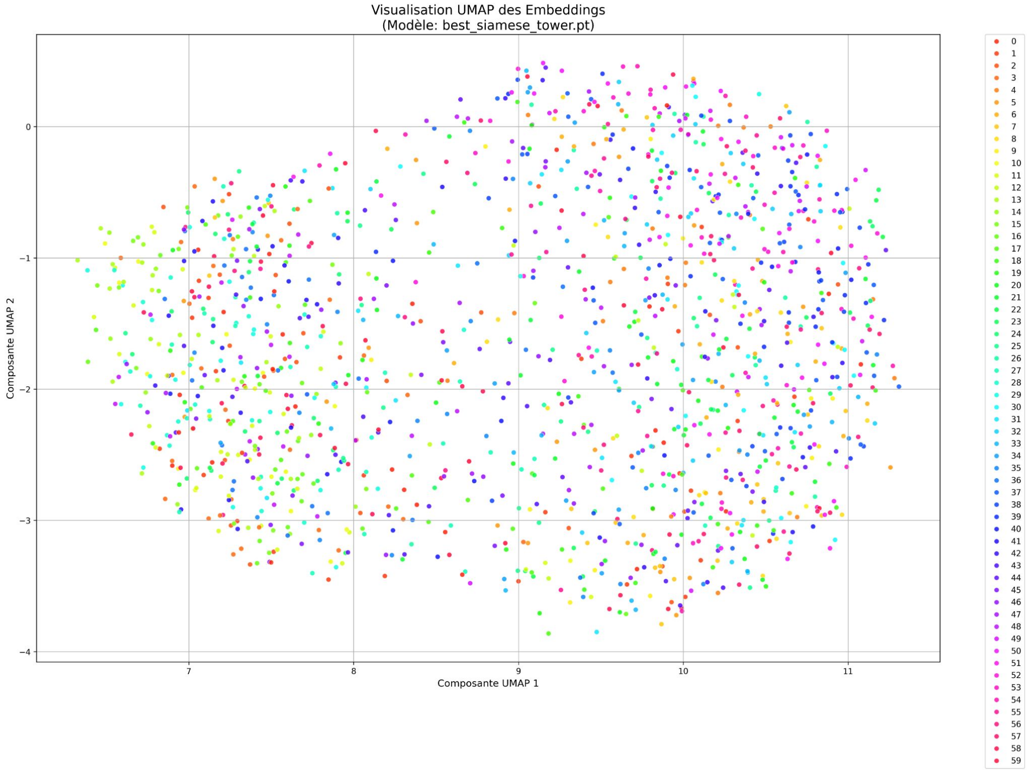
```
        Phase 1
percentage of labeled data from the initial unlabeled set: 18.75%
newly labelled: 89.67% (6068/6767 correct)
unlabeled: 0.00% (0/0 correctly unlabeled)
total labeled data: 89.67% (6068/6767 correct)

        Phase 2
percentage of labeled data from the initial unlabeled set: 25.70%
newly labelled: 77.98% (2511/3220 correct)
unlabeled: 25.21% (180/714 correctly unlabeled)
total labeled data: 86.76% (8045/9273 correct)

        Phase 3
percentage of labeled data from the initial unlabeled set: 29.49%
newly labelled: 73.44% (1850/2519 correct)
unlabeled: 28.24% (325/1151 correctly unlabeled)
total labeled data: 85.23% (9069/10641 correct)
```

# Observations

## Self-Supervised - NTU

- There is still challenge on the contrastive learning part, that the method still struggles to properly cluster the samples. Picture belows show visualisation of fc_256 layer responsible for the embedding during triplet loss.

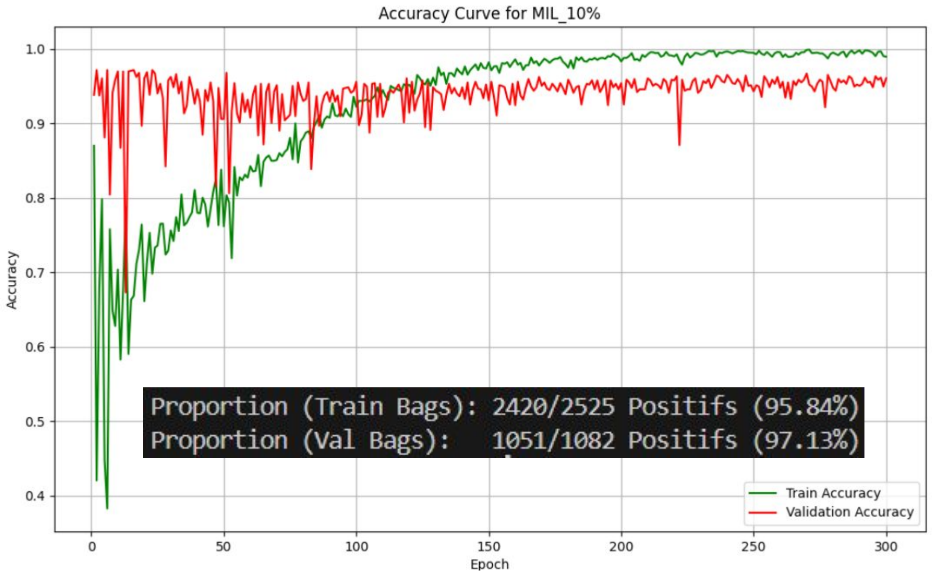


# Observations

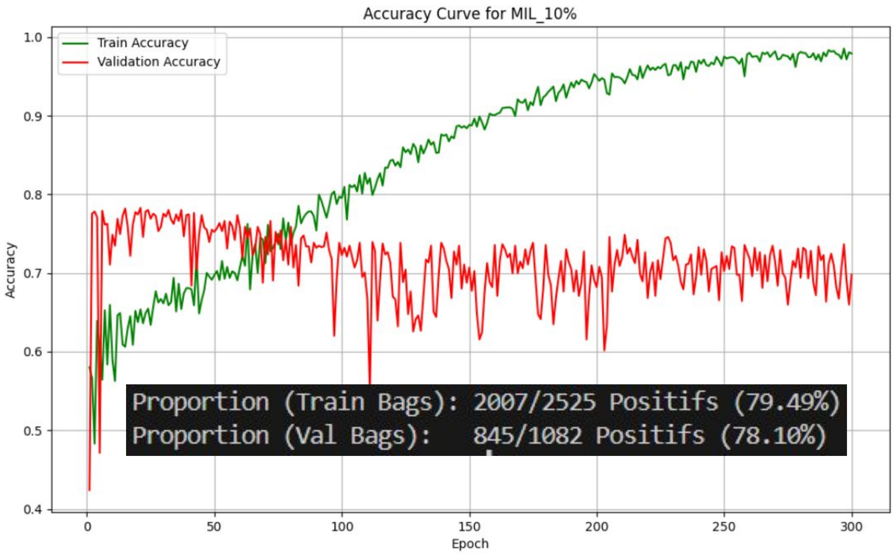
## Multiple instance learning

- In bagging, it is easier to classify certain class than the other ,which can be useful during this phase of training.
- Here, classifying two-person interaction agains others (daily action, medical condition) is easier, than distinguishing of daily actions with others (two-person, medical condition) - distinguising pattern is two or one person.




Positive : two-person + medical conditions

Negative : daily actions




Positive : two-person

Negative : daily actions + medical conditions

# Conclusion

- We show that it is possible to leverage unlabelled data to arrive on more accurate classification results, as opposed to only use available data.

**On NTU RGB+D:**

- **Weakly-Supervised :**
  → **Binary expert too weak** → introduces label noise → hurts multiclass performance.
- **Self-Supervised:**
  → **Contrastive model fails to form clusters** → poor feature representation.
- **Semi-Supervised:**
  → **Best performance**, especially in low-label regimes.
  → Scales with **training time** and **label confidence threshold**.

**On MocapLab:**

- **All methods outperform supervised baseline**.
- **Semi-Supervised:**
  → Strong improvement, but **sensitive to initial seed** and label quality.
- **Self-Supervised:**
  → **Better clustering** with contrastive methods.
- **Weakly-Supervised:**
  → Performance depends on **number of epochs** and model used.

# REFERENCES

## LITERATURE AND SOURCE CODES


[1] https://medium.com/swlh/multiple-instance-learning-c49bd21f5620

[2] https://medium.com/@aromalma/semi-supervised-learning-self-training-934a0c4f8700

[3] https://challengeenthusiast.com/self-supervised-learning-demonstratedg-f44ab8f2b45a

# ANY QUESTIONS ?

?

# THANK YOU

**HEDROUG Amine**

E-mail : amine.hedroug@telecom-sudparis.eu
Phone : 0665113444