

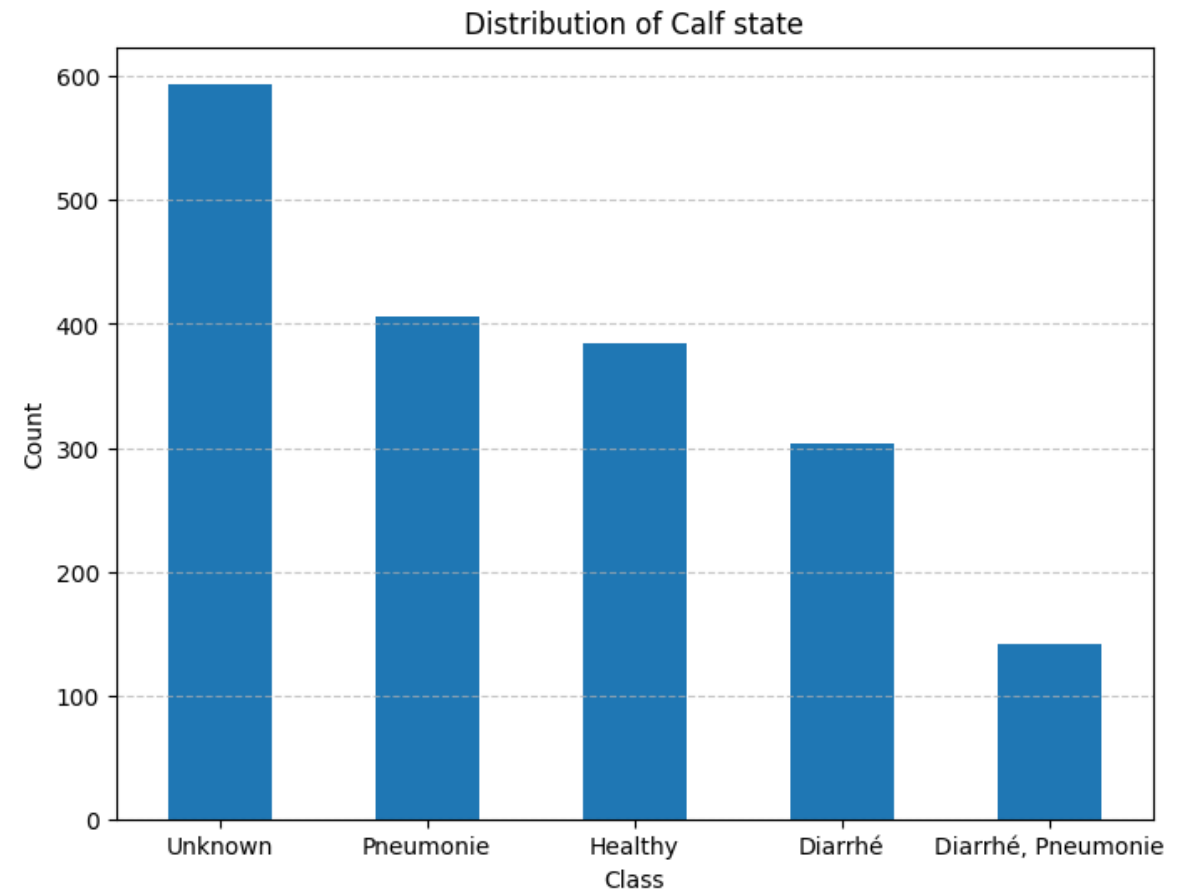
Extractions des images et
vidéos

- Les données videos

- Nous disposons de 9622 videos uniques d'une durée maximale de 1h.
- Ces videos couvrent 61 jours différents couvrant les mois de Fevriers, Mars et Avril 2023.

- Les données d'évaluations d'état de santé du veau

- Les données comportent 1829 lignes, couvrant 41 jours, avec en moyenne 44 évaluations par jour.
- La distribution des classes d'état de santé du veau est la suivante:



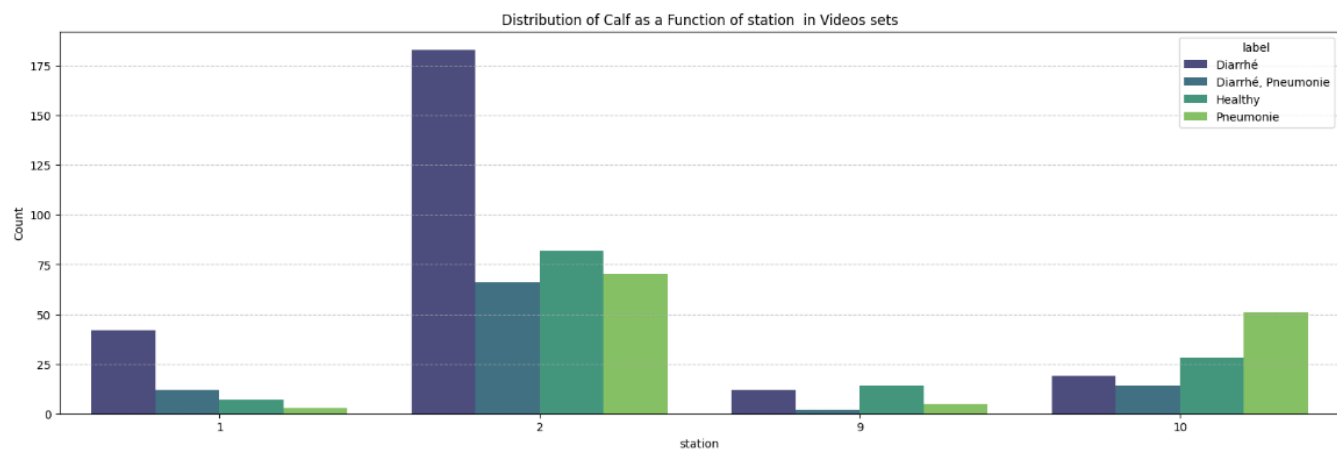
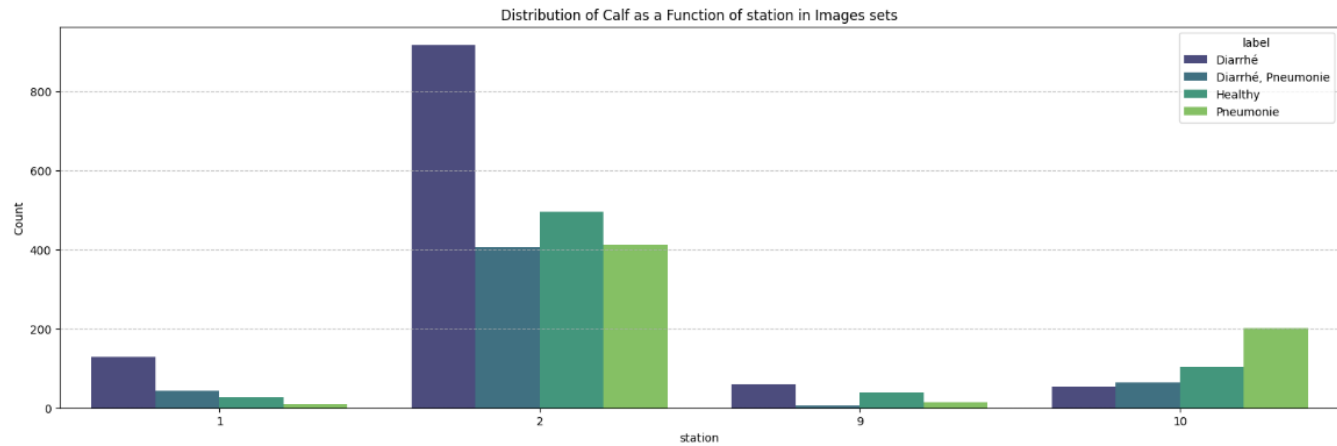
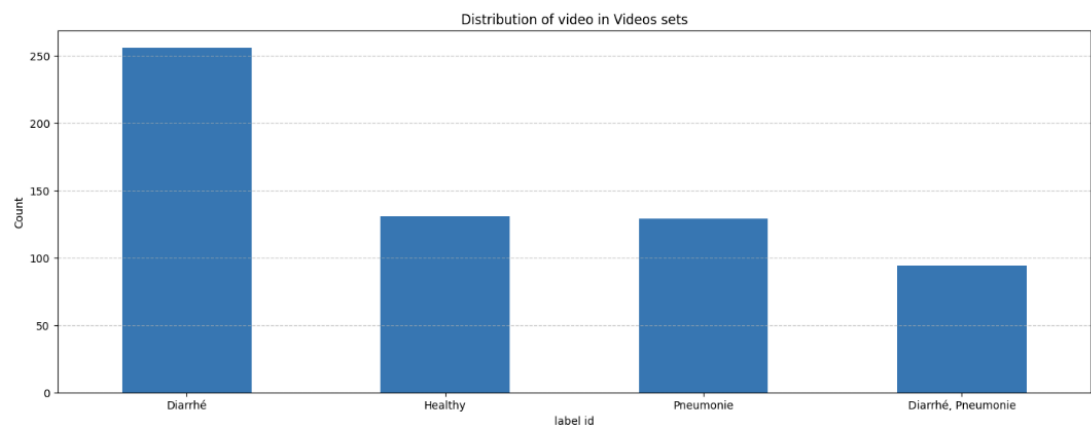
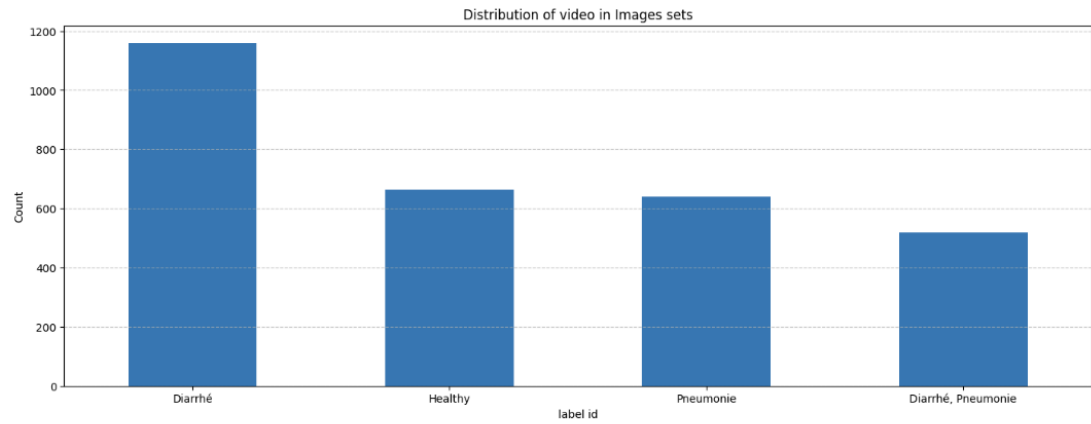
- Les données "Unknown" constituant 594 lignes (~32.47%), ne seront pas utiliser dans la suite du travail.

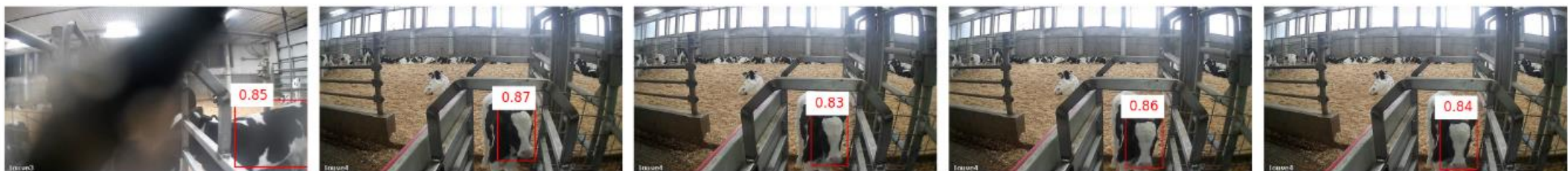
Les données de visites à la louve:

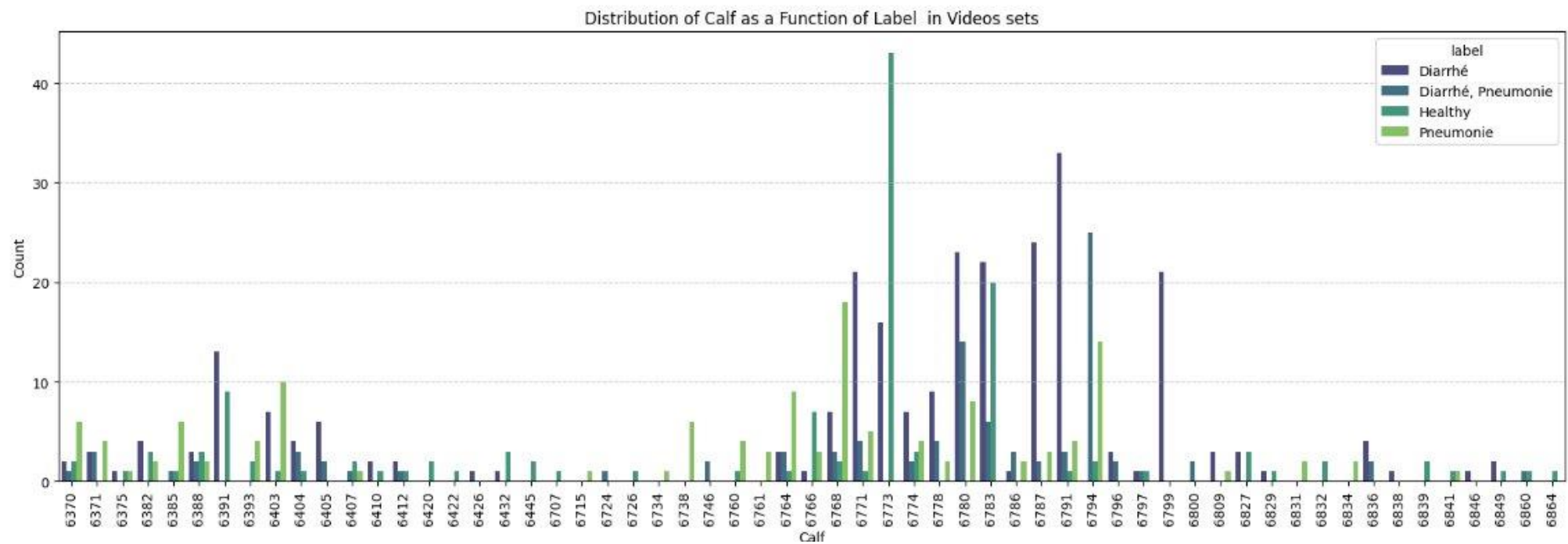
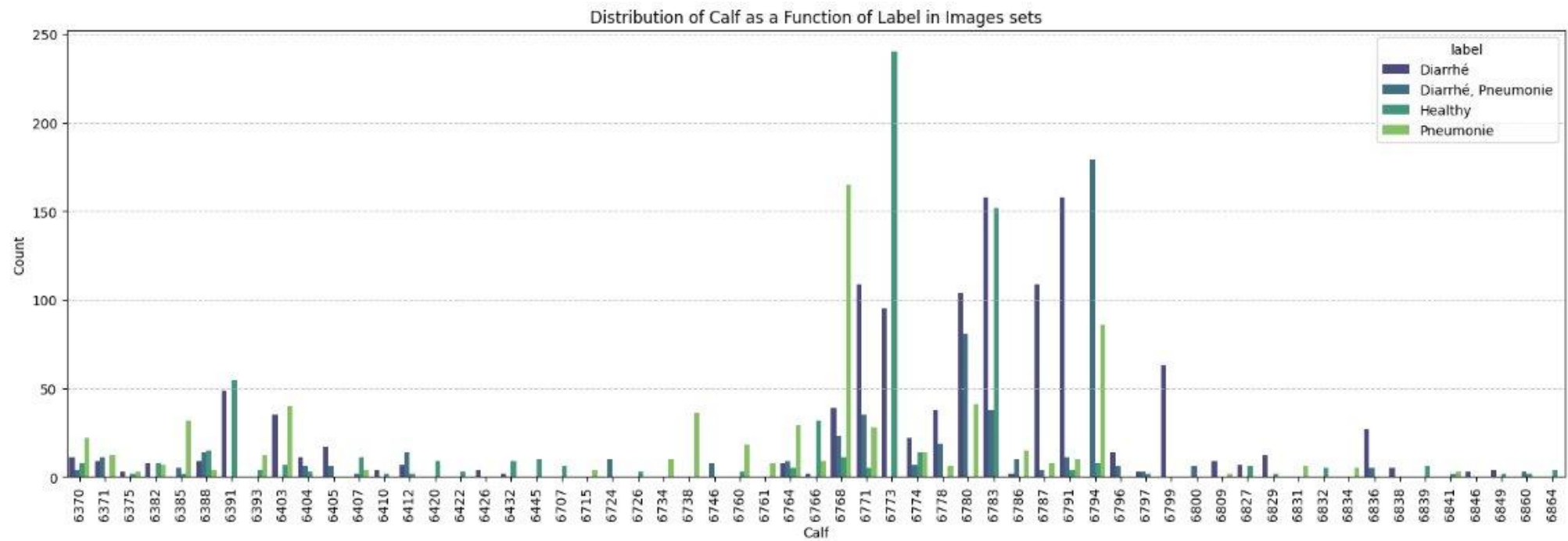
- Les données comportent 192755 lignes, dont seulement 23396 (~12.13%) ont eu lieu au cours des 41 jours au cours desquels l'etat du veau est connu.
- Parmi ces 23396 lignes, 18537 (~79.23%) ont une durée de plus de 0s.
- Parmi ces 18537 lignes, 15419 (~83.17%) appartiennent au channels 1, 2, 9, et 10.
- Parmi ces 15419 lignes, 14572 (~94.50%) ont eu lieu les jours dont on dispose des données videos.
- Parmi ces 14572 lignes, 13497 (~92.62%) sont des visites entre 6 et 22h de la journée.
- Parmi ces 13497 lignes, 13296 (~98.510%) ne concernent que les veaux dont l'état de sante a été évalué les jours de visites a la louve. Ces données représentent 6.89% des 192755 lignes initiales.
- On constate aussi certaines visites a la louve qui sont très proches de l'une de l'autre, avec une durée moyenne de 5 minutes entre les visites. Afin d'avoir des données permettant de distinguer le veau venant a la louve, les visites de la meme date et du meme channel avec un écart de moins de 3 minutes seront retirées.

- Au bout de ces processus, il reste 6362 lignes (~47.84%), parmi lesquelles nous utiliserons seulement 4380 lignes, pour les extractions de données. Ces 4380 lignes constituent les visites dont on connais l'etat de sante étant : 'Diarrhé', 'Pneumonie', 'Healthy', 'Diarrhé & Pneumonie'.

- En faisant une extraction d'image et videos sur les 4380 lignes, en utilisant un modèle yolo finetune sur la detection du visage de veau de face sur une séquence de 10s avant que le veau ne commence à manger à la louve, on obtient:
 - Avec le niveau de confiance du modèle yolo fixé a .80,
 - 610 videos, contenant au moins une frame sur laquelle est détecté un visage de veau
 - 2982 images avec annotation du visage de veau
 - Où on peut distinguer 59 veaux différents
- Les images suivantes montres des stats sur les données obtenues:







```
# total videos extracted, then total videos with a least one detection by y_face, and number of images extracted by y_face 3  
records[0].shape, records[0][records[0]['nfaces'] > 0].shape, records[1].shape
```

```
((96, 9), (29, 9), (118, 12))
```

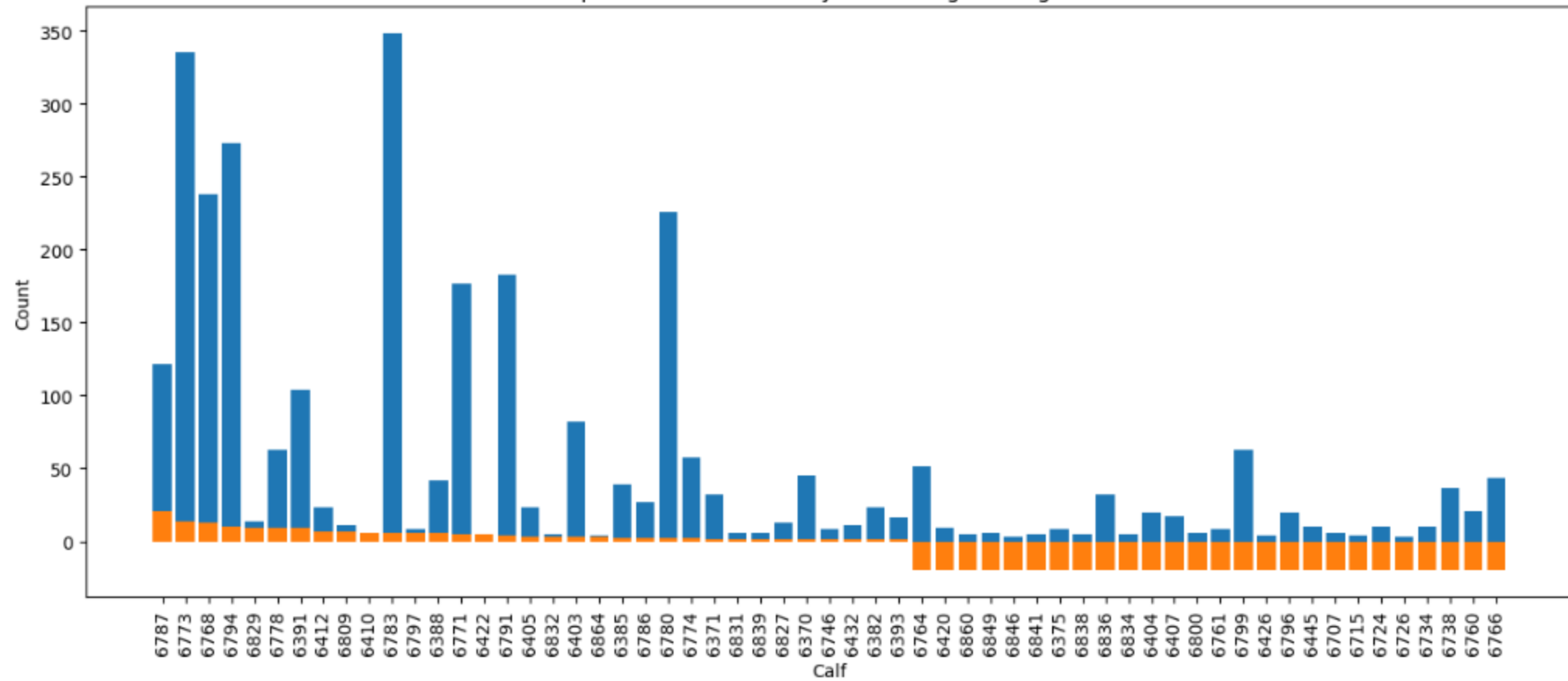
```
# total videos extracted, then total videos with a least one detection by y_face, and number of images extracted by y_face 4  
records[0].shape, records[0][records[0]['nfaces'] > 0].shape, records[1].shape
```

```
((96, 9), (28, 9), (70, 12))
```

```
# total videos with more than 1 detected with y_world, then total videos with a least one detection by y_face, and number of images extracted by y_face  
records[0].shape, records[0][records[0]['ncalfs'] > 0].shape, records[0][records[0]['nfaces'] > 0].shape, records[1].shape
```


```
((51, 10), (51, 10), (19, 10), (54, 12))
```

Graph with Calf Ordered by Decreasing Training Value



```
# conduct the Wilcoxon-Signed Rank Test
w = stats.wilcoxon(training_calf_distribution["count"], exported_calf_distribution["count"])
w.pvalue, w.pvalue < 0.05
```

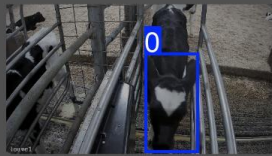
(1.0540120987485116e-06, True)

Transformations	P	R	mAP50	95					
hsv_h: 0.015, hsv_s: 0.7, hsv_v: 0.4, degrees: 0.0, translate: 0.1, scale: 0.5, shear: 0.0, perspective: 0.0, flipud: 0.0, fliplr: 0.5, bgr: 0.0, mosaic: 1.0, mixup: 0.0, copy_paste: 0.0, auto_augment: randaugment, erasing: 0.4, crop_fraction: 1.0	0.992	0.778	0.803	0.503					
hsv_h: 0.3, hsv_s: 0.3, hsv_v: 0.5, degrees: 40, translate: 0.3, scale: 0.3, shear: 0.5, perspective: 0.0, flipud: 0.0, fliplr: 0.5, mosaic: 1, mixup: 0.0	0.613	0.556	0.644	0.234					
degrees: 45, translate: 0.5, fliplr: 0.8, hsv_h: 0.3, hsv_s: 0.3, hsv_v: 0.5, scale: 0.3, shear: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.987	0.556	0.786	0.354					
degrees: 0, translate: 0.5, fliplr: 0.8, hsv_h: 0.3, hsv_s: 0.3, hsv_v: 0.5, scale: 0.3, shear: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.854	0.651	0.730	0.508					
degrees: 0, translate: 0.5, fliplr: 0.8, hsv_h: 0, hsv_s: 0, hsv_v: 0, scale: 0.3, shear: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.999	0.889	0.975	0.562	degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.5, shear: 0.5, hsv_h: 0.8, hsv_s: 0, hsv_v: 0, mosaic: 0.7, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.968	0.667	0.797	0.568
degrees: 0, translate: 0.5, fliplr: 0.8, hsv_h: 0.5, hsv_s: 0.5, hsv_v: 0.5, scale: 0.3, shear: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.942	0.667	0.700	0.432	degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.5, shear: 0.5, hsv_h: 0.8, hsv_s: 0.8, hsv_v: 0.8, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	1.000	0.537	0.741	0.403
degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.5, shear: 0, hsv_h: 0.5, hsv_s: 0.5, hsv_v: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.852	0.642	0.828	0.541	degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.7, shear: 0.5, hsv_h: 0.8, hsv_s: 0, hsv_v: 0, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.983	0.667	0.851	0.480
degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0, shear: 0.5, hsv_h: 0.5, hsv_s: 0.5, hsv_v: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.808	0.556	0.649	0.443					
degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.5, shear: 0.5, hsv_h: 0, hsv_s: 0, hsv_v: 0, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	1.000	0.655	0.705	0.425					
degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.5, shear: 0.5, hsv_h: 0, hsv_s: 0, hsv_v: 0.8, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	1.000	0.838	0.975	0.548					
degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.5, shear: 0.5, hsv_h: 0.8, hsv_s: 0, hsv_v: 0, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.968	0.667	0.797	0.568					
degrees: 0, translate: 0.5, fliplr: 0.8, scale: 0.5, shear: 0.5, hsv_h: 0.5, hsv_s: 0.5, hsv_v: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.672	0.686	0.750	0.507					
degrees: 45, translate: 0.5, fliplr: 0.8, scale: 0.5, shear:  hsv_h: 0.8, hsv_s: 0, hsv_v: 0, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.784	0.444	0.491	0.245					

46832 fcb9 f280228032022



0_1003025738263208__676



46832_fcb9_f280228032022



0103025738262208_676



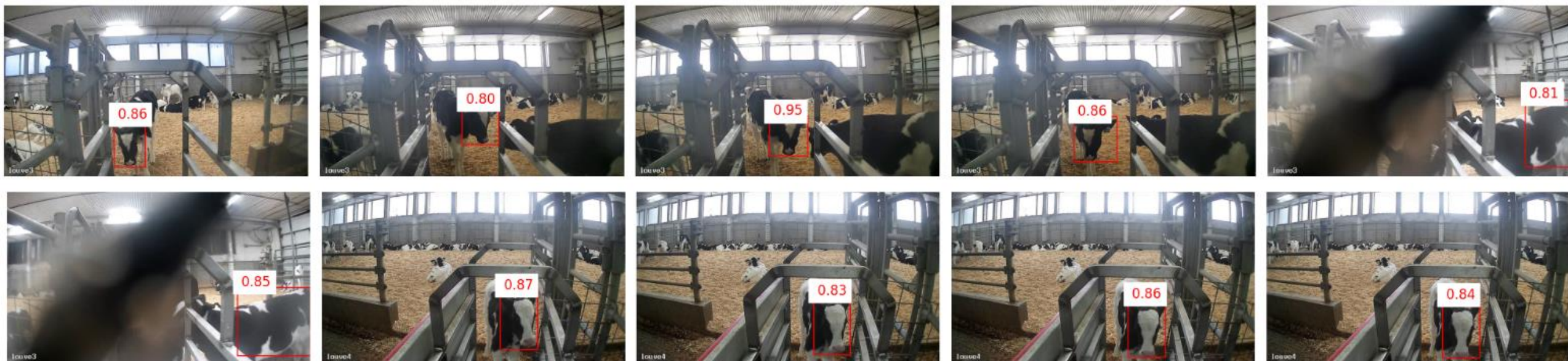
Yolo Face



Yolo World



Yolo Face



Yolo World



Yolo Face

Transformations	P	R	mAP50	95
hsv_h: 0.015, hsv_s: 0.7, hsv_v: 0.4, degrees: 0.0, translate: 0.1, scale: 0.5, shear: 0.0, perspective: 0.0, flipud: 0.0, fliplr: 0.5, bgr: 0.0, mosaic: 1.0, mixup: 0.0, copy_paste: 0.0, auto_augment: randaugment, erasing: 0.4, crop_fraction: 1.0	0.992	0.778	0.803	0.503

Yolo Face 2

degrees: 0, translate: 0.5, fliplr: 0.8, hsv_h: 0, hsv_s: 0, hsv_v: 0, scale: 0.3, shear: 0.5, mosaic: 1, mixup: 0.0, perspective: 0.0, flipud: 0.0	0.999	0.889	0.975	0.562
---	-------	-------	-------	-------

Yolo Face



Yolo Face 2



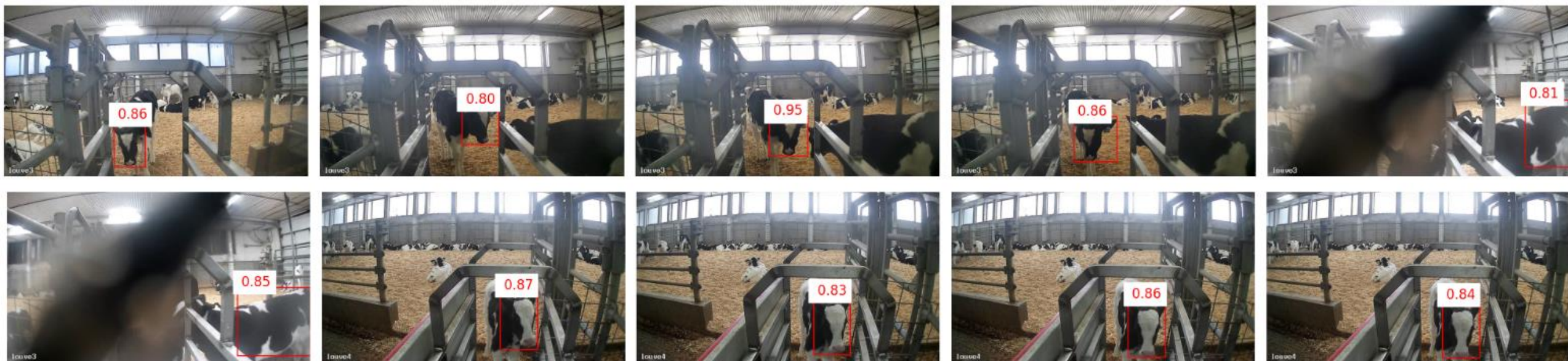
Yolo Face



Yolo Face 2



Yolo Face

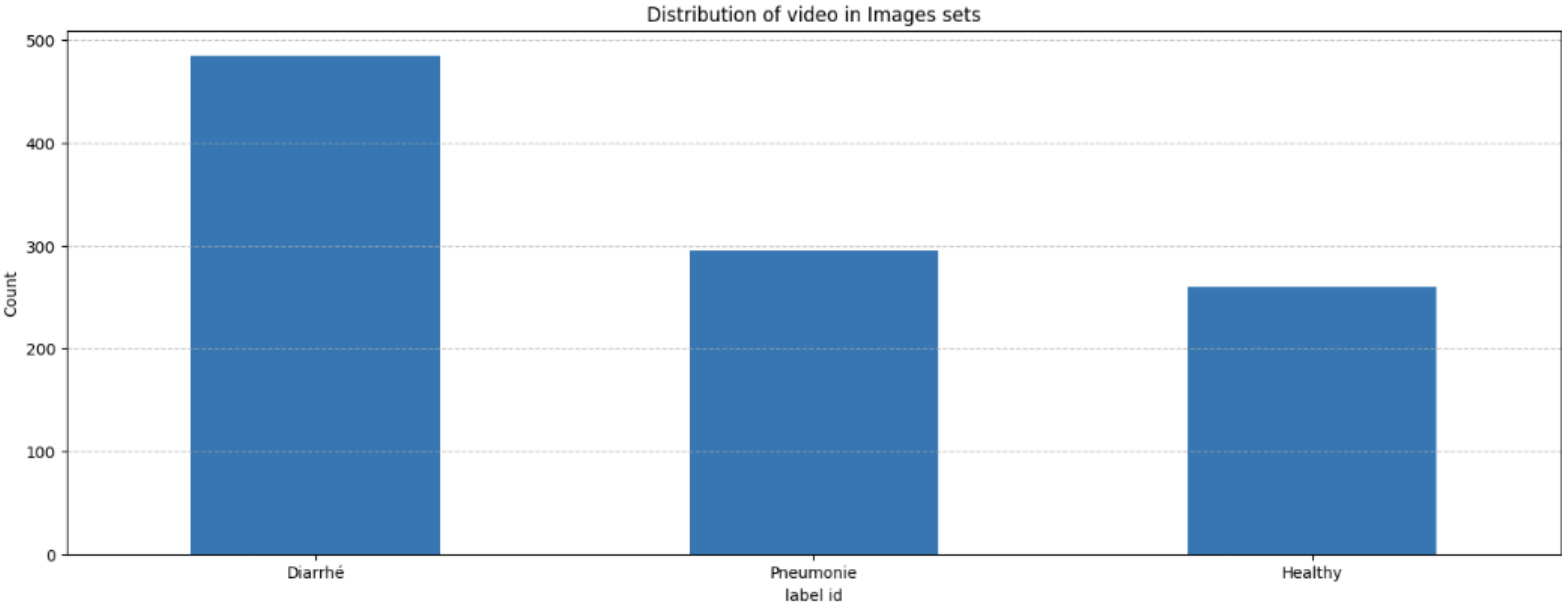


Yolo Face 2

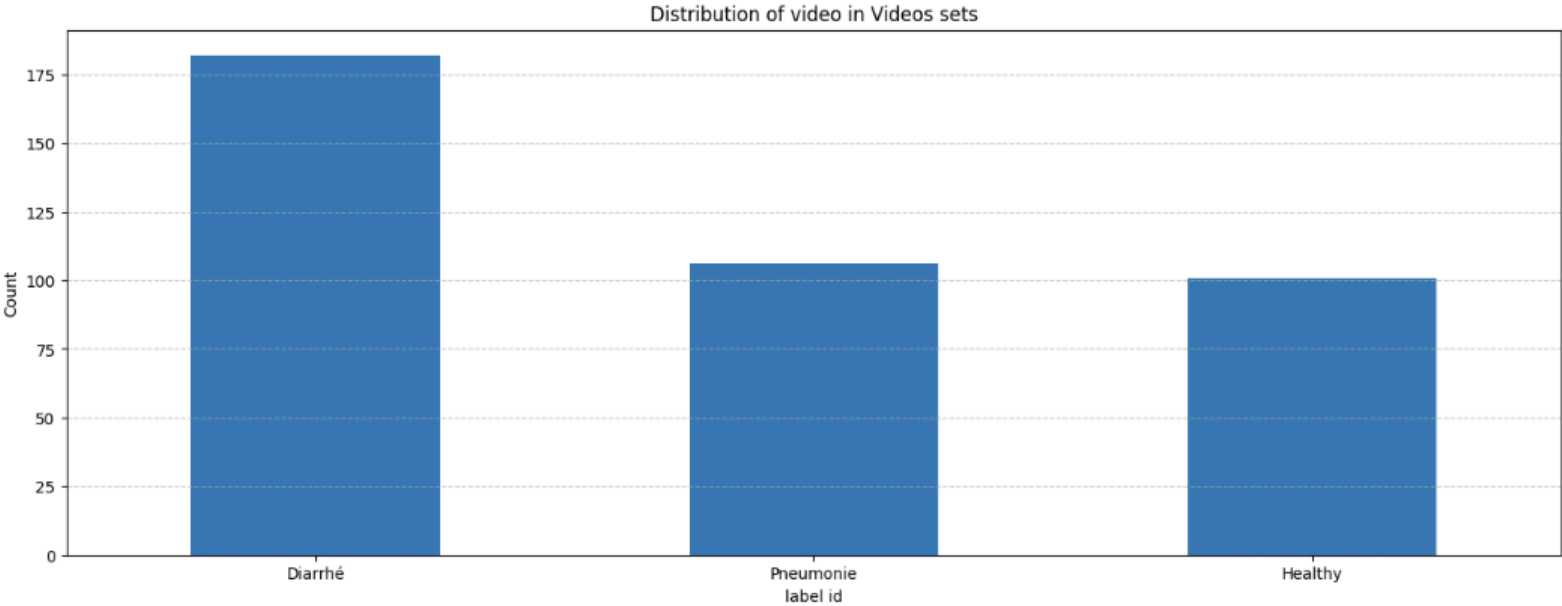


Yolo Face 2

1040 Images



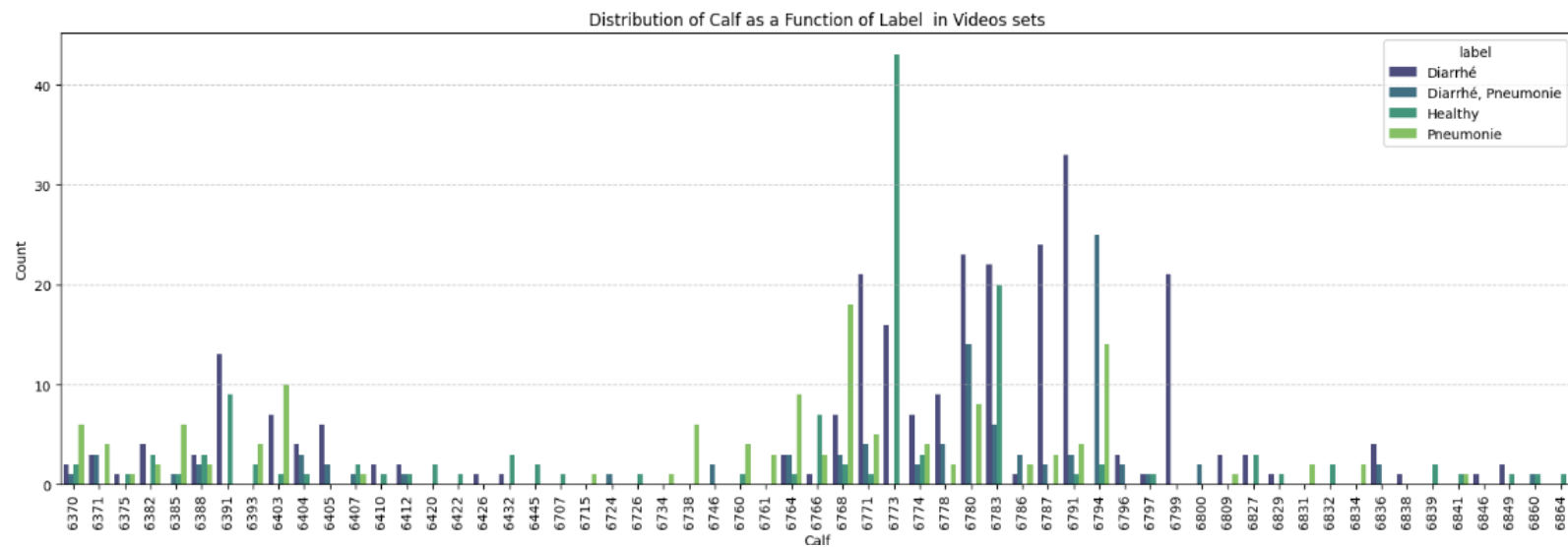
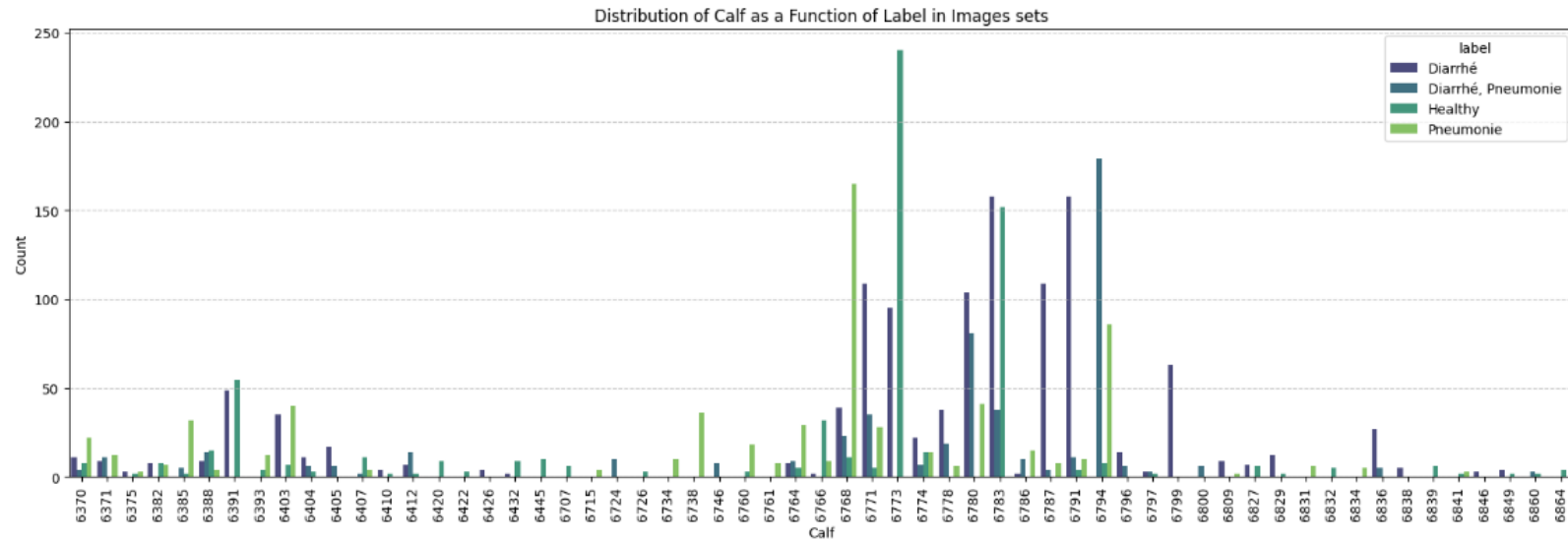
389 Videos



Yolo Face 2

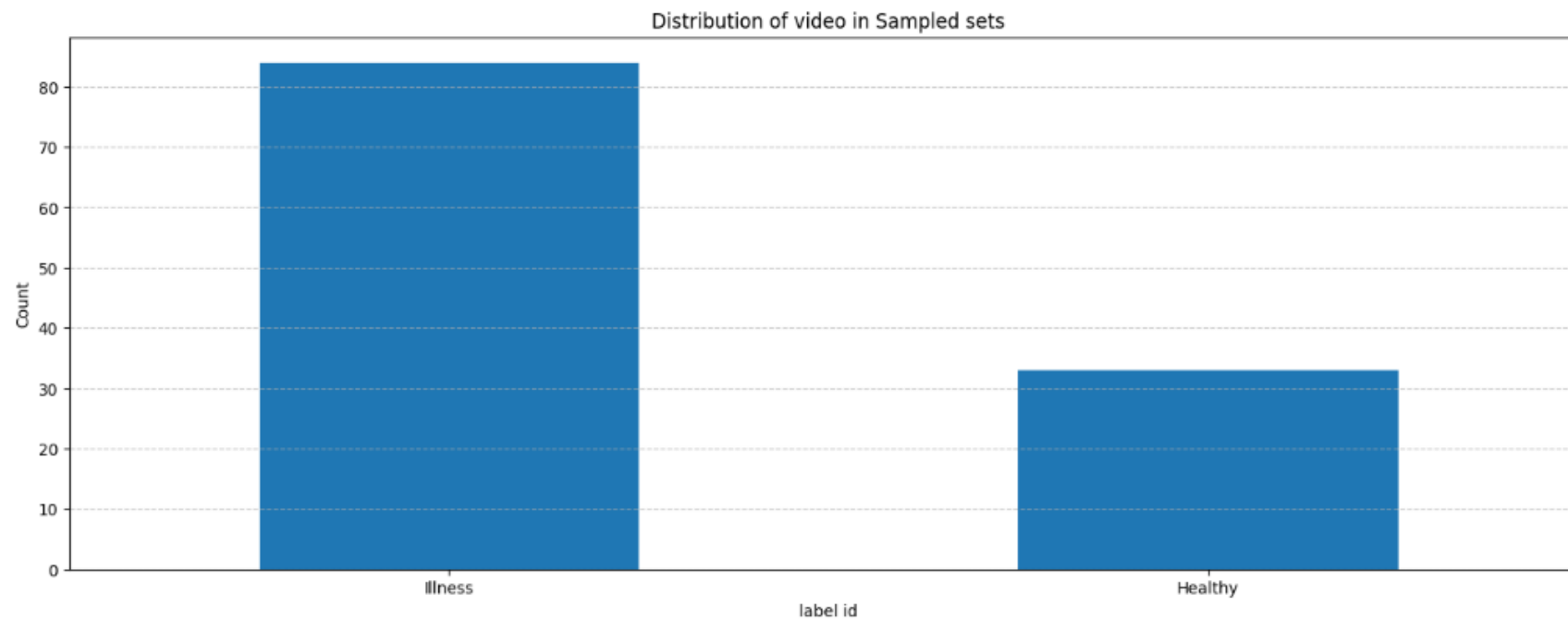
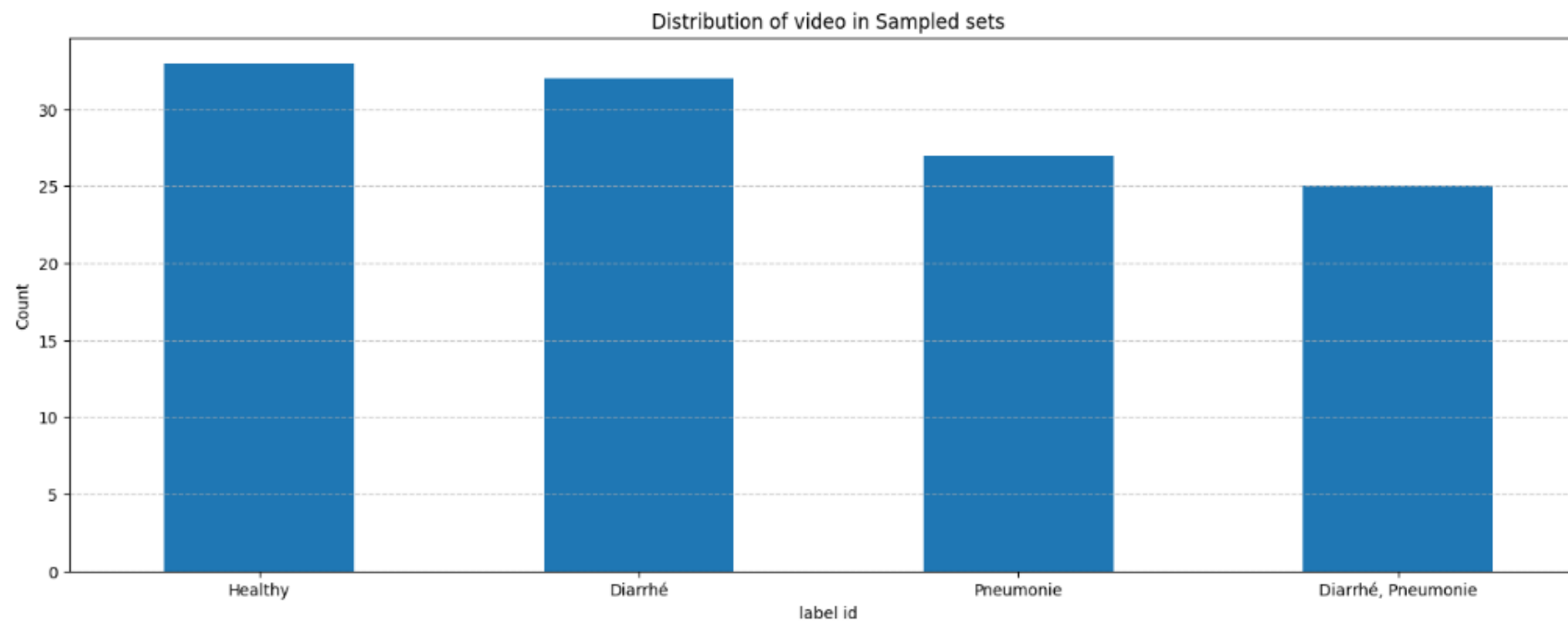


Best Training data

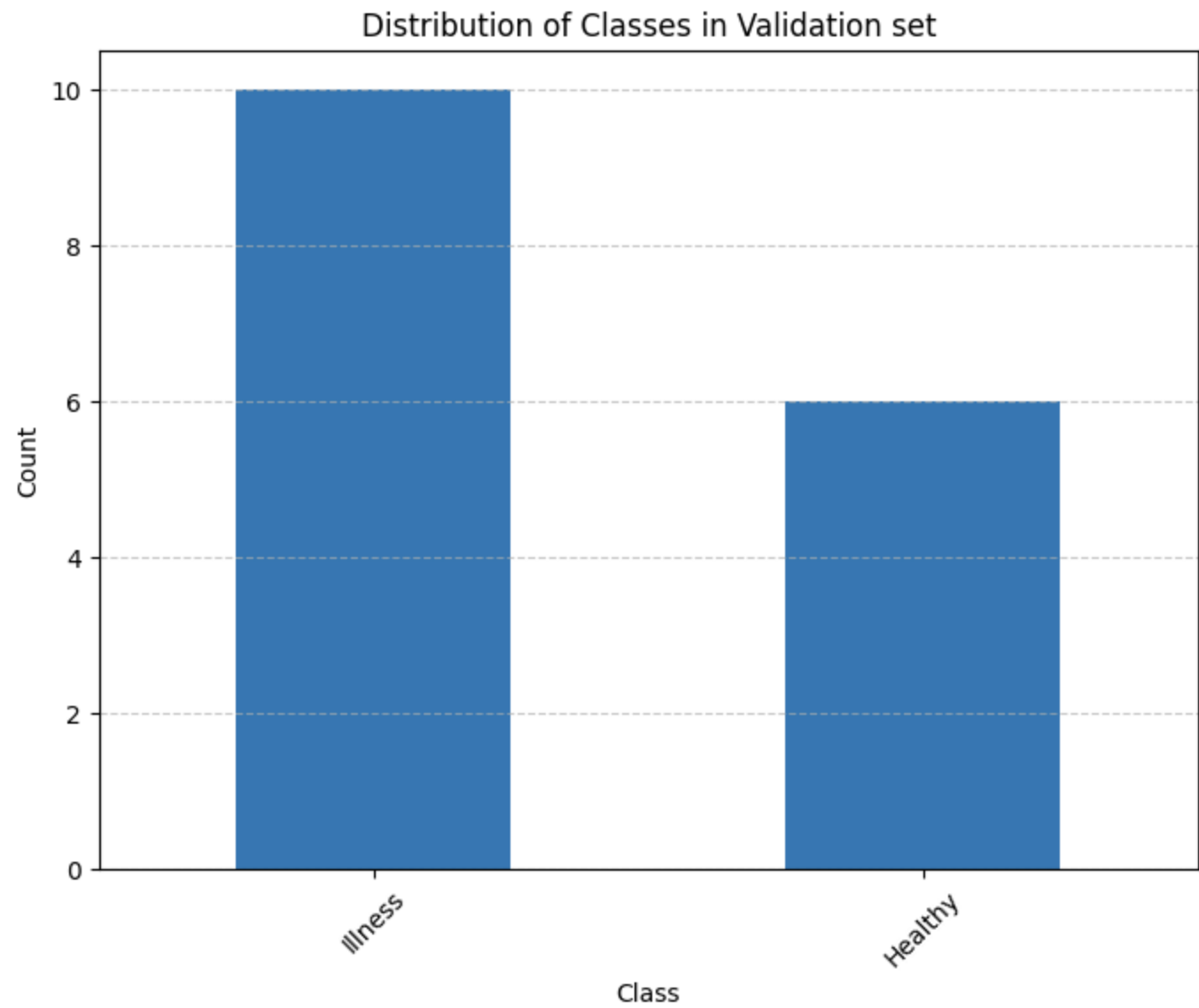


```
# Specify the columns to group by and the sample size
groupby_cols = ['calf', 'label']
sample_size = 2
# Get the uniformly sampled dataframe
sampled, not_sampled = uniform_sample_size_per_group(images_df, sample_size, groupby_cols)
sampled = sampled.sample(frac=1).reset_index(drop=True)
sampled.shape, not_sampled.shape
```

((254, 17), (2728, 17))



	Sample images	Whole images
Train	254	2338
Validation	2728	585
Test	16 et identique	



Modèle INTR

Training details

- Use similar set of data transformations to the one used to train Yolo (CIFAR10)
- Balance each batch
- Use a weighted loss
- Compute balanced accuracy
- Train over 10 epochs
- Use the best model base on lower loss on training

	Sample	Whole images	Previous performance s (for multi-class and without LOCO)
Accuracy en %	52.7778	58.3333	66.66
F1-score en %	56.6667	62.7778	66.46
Balanced Accuracy en %	59.4444	67.2222	-

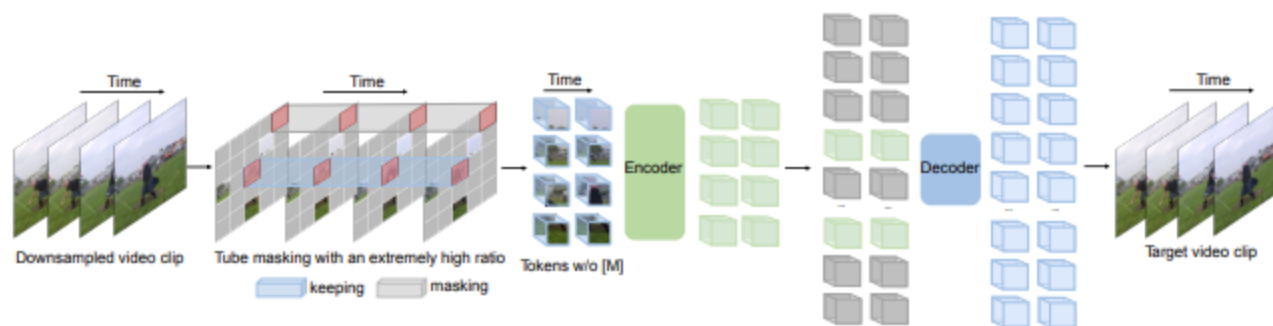


Figure 1: **VideoMAE** performs the task of masking random cubes and reconstructing the missing ones with an asymmetric encoder-decoder architecture. Due to high redundancy and temporal correlation in videos, we present the customized design of tube masking with an extremely high ratio (90% to 95%). This simple design enables us to create a more challenging and meaningful self-supervised task to make the learned representations capture more useful spatiotemporal structures.

VideoMAE

- Solution aux redondances entre les frames des vidéos de notre dataset
- Les performances sur de petits datasets

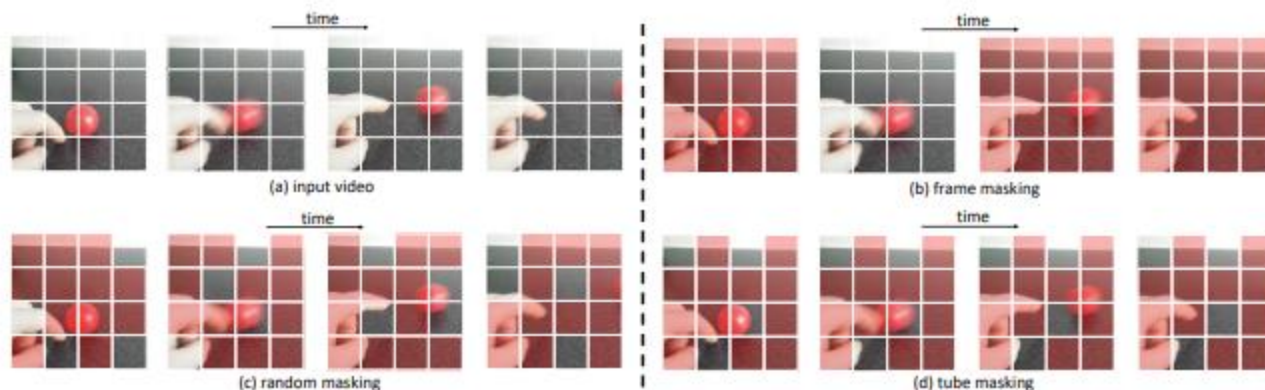


Figure 2: Slowness is a general prior in (a) video data [88]. This leads to two important characteristics in time: temporal redundancy and temporal correlation. Temporal redundancy makes it possible to recover pixels under an extremely high masking ratio. Temporal correlation leads to easily reconstruct the missing pixels by finding those corresponding patches in adjacent frames under plain (b) frame masking or (c) random masking. To avoid this simple task and encourage learning representative representation, we propose a (d) tube masking, where the masking map is the same for all frames.

Modèle VideoMAE

Training details

- Used a pretrained model on Kinetics dataset
- 10s of videos
- 16 frames per videos separate by 15 frames each
- Balance each batch
- Use a weighted loss
- Compute balanced accuracy
- Train over 10 epochs
- Use the best model base on lower loss on training

	Sample images	Whole images
Accuracy en %	40	40
F1-score en %	40	0
Balanced Accuracy en %	41.67	33.33

Modèle VideoMAE

Kinetics dataset

	Sample images	Whole images
Accuracy en %	40	40
F1-score en %	40	0
Balanced Accuracy en %	41.67	33.33

SSv2 dataset

	Sample images	Whole images
Accuracy en %	40	50
F1-score en %	25	44
Balanced Accuracy en %	45.83	50

Modèle VideoMAE

Kinetics dataset



(a) headbanging



(c) shaking hands



(e) robot dancing



(b) stretching leg



(d) tickling



(f) salsa dancing

SSv2 dataset

