Encadrants : Pr. ZAHARIA Titus, Doctorante OUENNICHE Kaouther

Étudiants : AUDY Quentin, BOULET Hugo, MAUBERT Jacques, RAKOTO David

Compte-rendu projet Cassiopée

Synthèse chronologique des travaux effectués

1. **Téléchargement du dépôt Git** sur la plateforme de Deep Learning.
   - Git source : https://github.com/kenshohara/3D-ResNets-PyTorch

2. **Téléchargement des bases de** données Kinetics 400 et UCF101, et préparation des données, en suivant les instructions du Git.

3. **Évaluation des modèles** R3D et R(2+1)D, préentrainés sur Kinetics 700, sur la base de validation de Kinetics 400.

4. **Création d'un programme** qui écrit un fichier **CSV** contenant les prédictions, les labels à prédire, et l'identifiant de la vidéo traîtée.

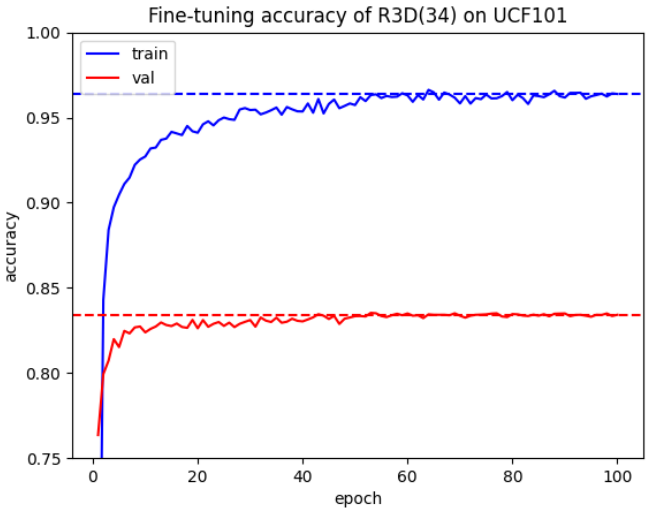
5. Recherche de la source des **mauvaises prédictions** des modèles.
   - Nous utilisons sur Kinetics 400 des modèles préentrainés sur Kinetics 700. Notre hypothèse est donc que les identifiants numériques des labels des vidéos de Kinetics 400 diffèrent de ceux de Kinetics 700, d'une manière qui nous échappe.
   - **Il est donc décidé de se focaliser plutôt sur UCF101**.

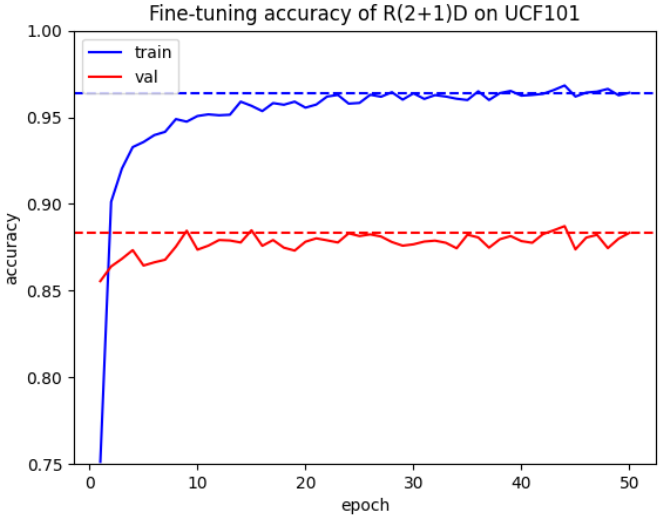
6. **Fine-tuning** de R3D(34), R3D(50), R(2+1)D(34) et R(2+1)D(50) sur UCF101.
   - Les **performances de nos modèles sont bonnes** (plus de 80% d'accuracy), mais elles sont **encore loin derrière les performances attendues**, surtout pour R(2+1)D.

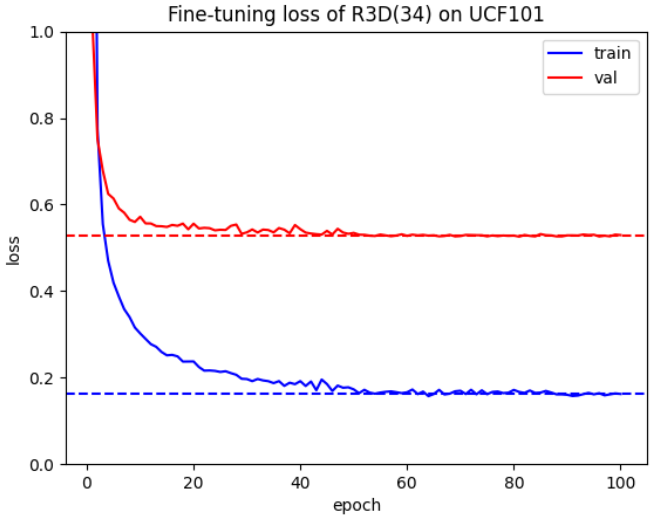
Modèle (Fine Tuning sur UCF-101)	Accuracy attendue	Accuracy obtenue
Modèle 3D (ResNet-34)	87,7	83,4
Modèle 3D (ResNet-50)	89,3	85,1
Modèle R(2+1)D (ResNet-34)	96,8	87,0
Modèle R(2+1)D (ResNet-50)	97,3	89,0

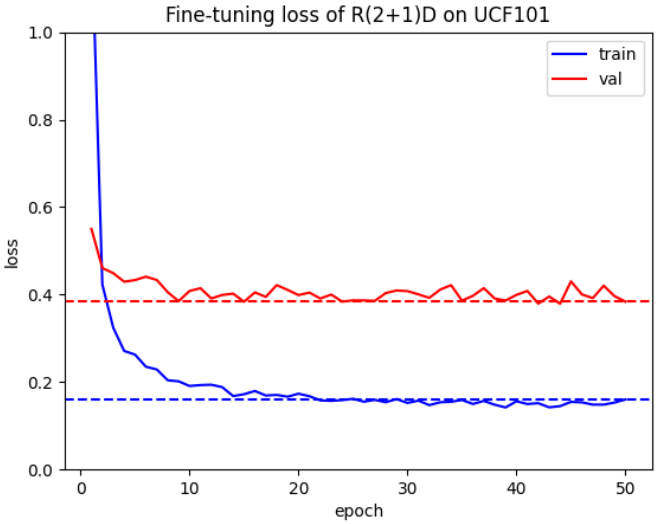
Résultats accuracy fine-tuning

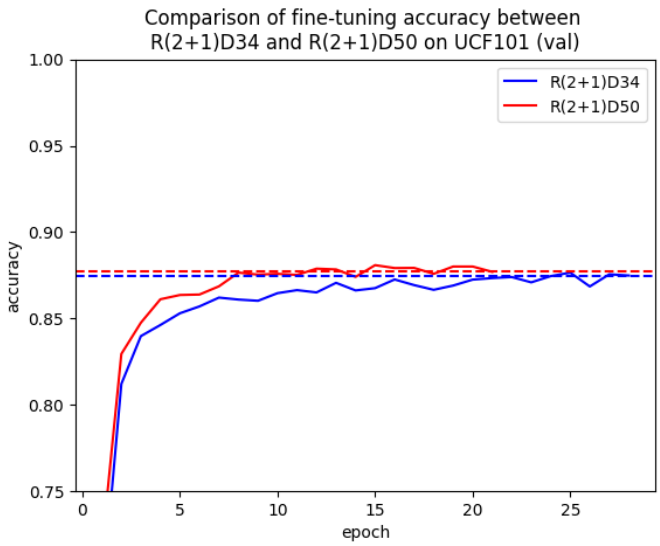
7. **Tracé de courbes** afin d'évaluer et comparer les performances du fine-tuning des différents modèles sur UCF101.
    - On remarque qu'on est loin d'overfit, donc il y a probablement de meilleurs choix d'hyperparamètres à faire, afin d'avoir de meilleures performances.

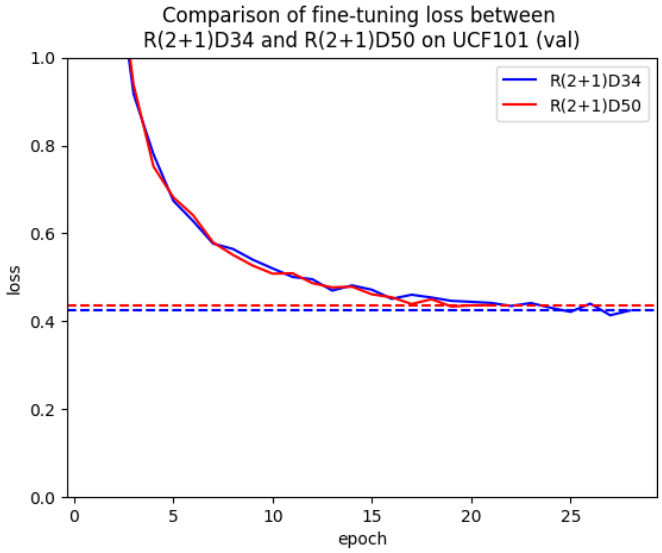




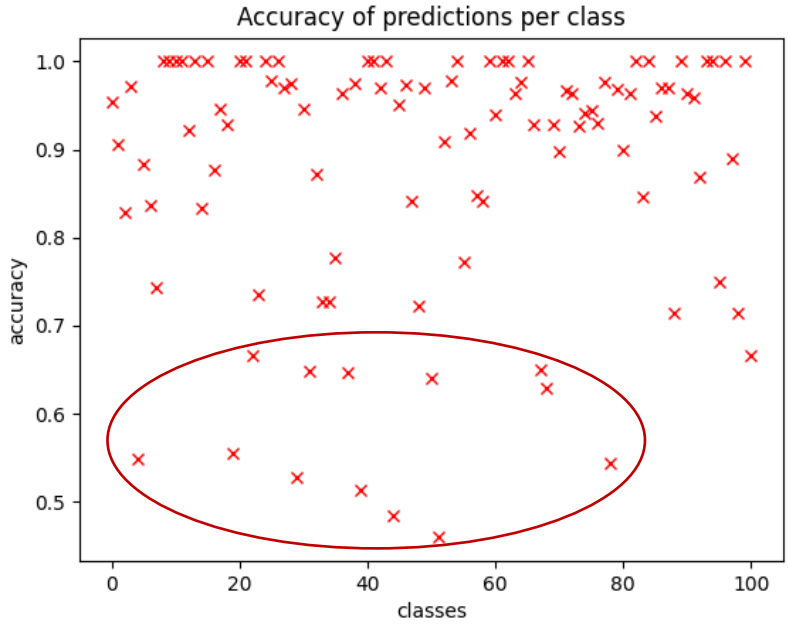








8. **Tracé** de **l'accuracy des prédictions pour chaque classe** de UCF101
   - On remarque que les prédictions sont particulièrement mauvaises pour certaines classes.



9. Création d'une **matrice de confusion**
   - La matrice, très volumineuse, est disponible dans l'archive sous le nom *matrice-de-confusion.xlsx*.

GROUND TRUTH/PRED	ApplyEyeMakeup	ApplyLipstick	Archery	BabyCrawling	BalanceBeam	BandMarching
ApplyEyeMakeup	42	1	0	0	0	0
ApplyLipstick	1	29	0	0	0	0
Archery	0	0	34	0	0	0
BabyCrawling	0	0	0	34	0	0
BalanceBeam	0	0	0	0	17	0
BandMarching	0	0	0	0	0	38
BaseballPitch	0	0	0	0	0	0
Basketball	0	0	0	0	0	0
BasketballDunk	0	0	0	0	0	0
BenchPress	0	0	0	0	0	0
Biking	0	0	0	0	0	0
Billiards	0	0	0	0	0	0
BlowDryHair	0	0	0	0	0	0
BlowingCandles	0	0	0	0	0	0
BodyWeightSquats	0	0	0	0	0	0
Bowling	0	0	0	0	0	0
BoxingPunchingBag	0	0	0	0	0	0
BoxingSpeedBag	0	0	0	0	0	0
BreastStroke	0	0	0	0	0	0
BrushingTeeth	0	2	0	0	0	0

Extrait de *matrice-de-confusion.xlsx*

10. **Synthèse de la matrice de confusion** sur les **10 classes aux pires prédictions**
    - Le fichier complet est disponible dans l'archive sous le nom *top-10-worst-class-predictions.txt*.
    - On remarque à ce moment que **les prédictions sont naturellement moins bonnes pour les classes qui peuvent être confondues avec d'autres**. Comme «marcher sur les mains » et « faire des pompes la tête en bas », « faire des fentes » et « faire des squats », ou encore « mettre du rouge à lèvre » et « se laver les dents ».

```
LongJump (50):
    LongJump (64.1%)
    JavelinThrow (15.38%)
    BandMarching (5.12%)
    Shotput (5.12%)
    CricketBowling (2.56%)
    BalanceBeam (2.56%)
    PoleVault (2.56%)
    ThrowDiscus (2.56%)

HandstandWalking (37):
    HandstandWalking (64.7%)
    HandstandPushups (14.7%)
    ParallelBars (8.82%)
    BoxingPunchingBag (5.88%)
    Basketball (2.94%)
    Fencing (2.94%)
```

Extrait de *top-10-worst-class-predictions.txt*

11. **Création d'une mini base de données**
    - La base de données créée est **disponible ici** : https://drive.google.com/drive/folders/1IQ00Fiu07sKuBCBD0AgUxr7ULz8MQnYM?usp=sharing (heureusement que le ridicule ne tue pas)
    - Elle contient **43 vidéos** de **maximum 6 secondes**, représentant au total **27 actions différentes**, présentes dans la base de UCF101. Les actions choisies sont des **actions que la modèle est supposé avoir du mal à prédire**, ou simplement des actions qui étaient filmables à portée de main.
    - Parfois ce sont de **vraies actions** qui ont été filmée (faire des pompes, faire des squats, se laver les dents, mettre du maquillage, etc…), et parfois ce sont de simples **imitations des actions** (se couper les cheveux (ciseaux faits avec les doigts), faire un dunk, mettre un coup de poing, etc…).
    - Parfois, la même action est filmée, mais sous plusieurs angles différents.

Capture d'écran de quelques vidéos

12. Prédictions sur notre base de données

- Nous avons utilisé **R(2+1)D(50), préentraîné sur Kinetics 700**, que nous avons **réentraîné sur UCF101**, pour prédire les labels des différentes vidéos.
- Le **fichier contenant toutes les prédictions obtenues**, ainsi que les notes sur chacune, est **disponible dans l'archive** sous le nom *resultats-predictions-propre-bdd.xlsx*.
- **Nous obtenons une accuracy de 58% sur notre base de données**, composées spécifiquement en partie de vidéos où on sait que les prédictions sont mauvaises (et certaines vidéos difficilement prédisables dans l'absolu)

Correct ?	Prediction	Ground truth	Note
OUI	ApplyEyeMakeup		Amélie
OUI	ApplyEyeMakeup		David
NON	BrushingTeeth	ApplyLipstick	Amélie, confusion classique entre les deux actions semblables
OUI	ApplyLipstick		David
NON	SoccerJuggling	Basketball	Le modèle voit un ballon mais a du mal à interpréter l'action
NON	WritingOnBoard	BasketballDunk	Sans trop de surprise
OUI	Billiards		Billard anglais, alors que modèle entraîné sur billard américain
OUI	BodyWeightSquats		De face
NON	WallPushups	BodyWeightSquats	De profil
NON	WallPushups	BoxingPunchingBag	Sans surprise
OUI	BrushingTeeth		
OUI	CuttingInKitchen		
OUI	Haircut		Ciseau avec les doigts, quasi que femmes dans bdd

Extrait de *resultats-predictions-propre-bdd.xlsx*

- Il est intéressant de remarquer que parfois, même quand on imite une action, le modèle reconnaît l'action imitée.

- On remarque notamment que les vidéos de certaines classes sont peu diversifiées. Exemple : les vidéos de l'action « taper au clavier » sont tous des plans très rapprochés où on ne voit que les mains écrire. Donc dès qu'on veut faire une prédiction d'une personne qui tape sur un clavier sur un plan plus large, le modèle peine à prédire le bon label.

13. Prédictions top 3 et top 5

- Nous avons refait les prédictions sur notre base de données, **en notant cette fois si le ground truth est dans le top 3 ou dans le top 5 des classes prédites**.
- Le **fichier contenant toutes les prédictions obtenues** est **disponible dans l'archive** sous le nom *resultats_topk.xlsx.*

Correct ?	Top 3 ?	Top 5 ?	Ground truth
OUI	OUI	OUI	ApplyEyeMakeup
NON	OUI	OUI	ApplyEyeMakeup
NON	OUI	OUI	ApplyLipstick
OUI	OUI	OUI	ApplyLipstick
NON	OUI	OUI	Basketball
NON	NON	NON	BasketballDunk
OUI	OUI	OUI	Billiards
OUI	OUI	OUI	BodyWeightSquats
OUI	OUI	OUI	BodyWeightSquats
NON	NON	OUI	BoxingPunchingBag
OUI	OUI	OUI	BrushingTeeth
OUI	OUI	OUI	CuttingInKitchen
OUI	OUI	OUI	Haircut
NON	OUI	OUI	Haircut

Extrait de *resultats_topk.xlsx*

- On remarque finalement que les seules **vidéos pour lesquelles la ground truth n'est pas dans le top 5 des labels prédits par le modèle**, sont les **vidéos impertinentes**, comme le dunk sur un panier miniature, le coup de poing au ralenti (pour « Punch »), ou encore l'imitation de combat de sumo. Les seules exceptions sont les **vidéos de YoYo**, pourtant semblables à celles présentes dans la base de données (bien que je sois moins bon que les gens qui s'y trouvent). Peut-être que le modèle à du mal à identifier le yoyo (l'objet) sur l'image.
  - ➔ **Le ground truth se trouve dans le top 5 des labels prédits pour 84% des vidéos de notre base.**
- Autrement, sur les vidéos qui restent pertinentes, **lorsque la prédiction est mauvaise, le ground truth se retrouve souvent dans le top 3 des labels prédits**.
  - ➔ **Le ground truth se trouve dans le top 3 des labels prédits pour 81% des vidéos de notre base.**

Encore merci de nous avoir accompagnés sur ce projet.