

## Étape 2

Amélioration des descripteurs, ou création de nouveaux, en utilisant  
de l'apprentissage

Les groupes sont formés sous moodle

Tous les compte-rendus et la présentation sont à rendre par groupe  
(c'est-à-dire, 1 seul dépôt suffit)

Les connaissances seront évaluées individuellement

Thibault BLANC-BEYNE, Axel CARLIER, Sylvie CHAMBON, Thierry MALON

11 janvier 2021

## Page destinée aux tuteurs

### Éléments pouvant intervenir dans les discussions / questions / réponses

- Méthodologie de l'apprentissage : ensemble d'apprentissage, de validation, voire de test
- Ajout de sémantique "apprise" dans les descripteurs ; utilisation par exemple de l'activation de couches intermédiaires de réseaux pré-entraînés sur de larges bases (type ImageNet)
- Eventuelle réutilisation de descripteurs mis au point en 1e étape, avec intégration de l'apprentissage (pas forcément réseaux de neurones, certains étudiants ont évoqué SVM par exemple) pour mettre au point certains paramètres
- Utilisation d'auto-encodeurs (voire même certaines variantes, comme les auto-encodeurs épars) pour mettre au point un espace latent de descripteur
- Utilisation de réseaux siamois pour mettre au point l'espace de descripteur (approches type DeepFace ou FaceNet)

### Approches qui devraient être lues/étudiées

- Auto-encodeur (mais déjà vu en cours)
- Approches similaires à la détection de visage : DeepFace, FaceNet (présenté très rapidement en cours)
- *CutOut*, une technique d'augmentation de données qui permet de limiter le sur-apprentissage et d'établir des descripteurs plus robustes aux occultations

### Difficultés/questions que les étudiants doivent absolument identifier pour faire les bons choix

- Comme dans l'étape 1, le pré-traitement des images sera important (normalisation, mise à l'échelle, etc.)
- La question de la division des bases de données en ensembles d'apprentissage/validation/test nécessite réflexion : il est important de répartir les données le mieux possible pour éviter d'avoir une sur-représentation ou une sous-représentation de certaines catégories dans tel ou tel ensemble. Un exemple pour les Pokemon pourrait être de s'assurer d'une répartition des "types" (eau, feu, etc.) à peu près homogène dans les différents ensembles.
- Pour le cas d'utilisation Pokemon, les étudiants devraient très clairement se heurter au problème du nombre limité de données. Des pistes pour s'en sortir seraient d'utiliser l'augmentation de données (**Attention à appliquer des transformations qui sont pertinentes vis à vis de la base de données**), ou même pourquoi pas d'aller chercher des données similaires (*sprites* issues d'autres jeux par ex) ce qui dans le cas des auto-encodeurs permettra d'établir un espace latent plus pertinent.
- Si les étudiants se tournent vers les réseaux siamois, une des difficultés est d'établir des paires (ou des triplets) d'entraînement qui apportent une information la plus pertinente possible. C'est un élément essentiel du papier FaceNet (section 3.2 *Triplet Selection*).

### Approches naturellement envisageables pour les étudiants

- Approche hybride combinant des descripteurs réalisés en 1e étape avec des descripteurs portant une information plus sémantique vus dans cette étape.
- Certains pourraient être tentés d'aller vers les GANs... Leur rappeler la difficulté d'entraîner ces GANs, qu'ils ont expérimenté lors du TP d'IAM correspondant.
- Il faut que les étudiants prennent le temps de comparer les approches deep learning avec les approches de l'étape 1 !

## Objectifs d'apprentissage visés par cette étape

- Mettre en oeuvre l'entraînement d'un réseau de neurones du début à la fin
- Connaître une ou plusieurs approches de descripteurs basés apprentissage profond
- Connaître les forces et faiblesses comparées des approches vision et des approches apprentissage profond

## Étape 2

### Reconnaissance en utilisant des descripteurs issus de l'apprentissage



#### Temps alloué pour cette étape

- **Aller** : 12/01 à 13h45 : **Compte rendu noté**
- **Autonomie/travail individuel** : 14/01 à 8h, 18/01 à 10h
- **Intermédiaire** : 21/01 à 10h : **Compte rendu noté**
- **Autonomie/travail individuel** : 25/01 à 10h, 26/01 à 15h45, 29/01 à 10h
- **Retour** : 01/02 à 10h : **Compte rendu noté**
- **Mise en commun** : 04/02 à 8h : **Présentation notée + évaluations à faire seul**

Tout travail rendu sans être présent à la séance consacrée entraînera une note à 0 pour la personne absente, sauf justification de l'absence.

#### Mission

Votre employeur arrive un matin dans votre bureau en déclarant que pour des raisons *Marketing*, **il est maintenant extrêmement important d'afficher que l'entreprise fait de l'Intelligence Artificielle.**

Il vous est donc demandé de revoir votre copie et d'utiliser l'apprentissage profond pour établir des descripteurs nécessaires à la reconnaissance.

Comme vous avez été bien formé, vous savez que l'apprentissage profond n'est pas la panacée et que, peut-être, une approche pertinente utilisera des descripteurs issus de l'apprentissage mais **en conjonction ou en complément** des approches vues lors de la première étape.

Votre objectif dans cette nouvelle étape est non seulement de mettre au point une ou plusieurs méthodes de reconnaissance qui se basera sur des descripteurs appris grâce à des réseaux de neurones, mais aussi de mettre les performances de ces méthodes en perspective avec vos résultats précédents, voire même de combiner les approches si vous en avez le temps et si vous le jugez pertinent.

**Toutes les solutions que vous proposerez doivent être testées de manière pertinente afin de déterminer leurs limites.**

#### Livrables

A l'issue de cette seconde étape, il vous est demandé de déposer sur Moodle : une présentation de la réponse à cette question. Nous vous donnerons plus de détails en temps voulu sur les éléments à faire apparaître dans cette présentation.

Pendant la mise en commun, en s'appuyant sur cette présentation, il faudra :

- d'une part, expliquer l'algorithme proposé et, d'autre part, faire une démonstration du code développé.
- d'autre part, comparez les forces et faiblesses des approches retenues dans les 2 étapes, en vous appuyant à la fois sur des données quantitatives mais également sur des exemples visuels précis.

## Références

Vous trouverez sous Moodle les diapositives du cours sur les auto-encodeurs que vous avez déjà suivi. L'espace latent appris par les auto-encodeurs est un bon candidat pour servir d'espace de descripteur.

Nous vous fournissons également deux articles de recherche, présentés très brièvement en cours, sur la reconnaissance de visages : DeepFace et FaceNet, qui utilisent tous les deux des réseaux siamois. La reconnaissance de visage est un problème très similaire à la réidentification...

Un dernier article vous sera utile si vous avez envie de mettre en place d'augmentation de données : il s'agit de *CutOut*.