

## **IA : Computer Vision, traitements, analyses & génération d'images/vidéos**

*Devenir un expert de la vision par ordinateur*

### DESCRIPTION

Les données non structurées issues de l'image sont omniprésentes en industrie et dans notre quotidien (face id, radars, détection d'objets, d'anomalies, caméras des chaînes de production, contrôle de qualité, médecine...).

Pour autant leur exploitation par les organisations s'avère complexe sans une approche réfléchie et structurante pour en tirer le meilleur parti. Les progrès en deep learning ont permis de rendre plus accessible l'exploitation des données images et vidéos.

Cette formation prépare les data scientists à maîtriser les données images et vidéos, au sein d'un contexte technologique innovant et en particulier au cours d'un projet d'Intelligence Artificielle.

A travers des exercices, vous apprendrez à structurer et créer des modèles de machine learning sur ces données. A la fin de la session, vous disposerez d'une compréhension solide du potentiel et de l'état de l'art en Computer Vision. Les mises en pratique vous permettront d'être indépendant pour déployer et créer de la valeur sur ces données, et devenir un expert de la vision.

### OBJECTIFS PEDAGOGIQUES

- Découvrir et manipuler la donnée image et vidéo
- S'approprier le fonctionnement des modèles Convolutional Neural Network (CNN)
- Implémenter un modèle CNN en classification, détection/segmentation, connaître l'état de l'art
- Évaluer ses modèles et choisir les bonnes métriques
- Connaître les méthodes pour interpréter les modèles CNN et leurs prédictions : IA de confiance (AI Act)
- S'approprier les méthodes pour faire des inférences avec peu d'images (petite base de données)
- Appréhender le fonctionnement de DALL-E 2 et des modèles de diffusion
- Savoir générer de la donnée image ou modifier une image (améliorer la qualité de l'image, débruiter)

**Stage pratique**  
Data Science

Code :  
**CVTAG**

Durée :  
**3 jour(s) (21,00 heures)**

Exposés : **45 %**  
Cas pratiques : **45 %**  
Echanges d'expérience : **10 %**

**Inter-entreprises :**  
Prochaines sessions  
disponibles [sur notre site web](#).  
Tarif : 2 500,00 € HT /  
participant

**Intra-entreprise :**  
Tarifs et dates sur demande.

- Optimiser et embarquer ce genre de modèle (sur un raspberry)
- Posséder une cartographie des modèles à utiliser pour faire de l'inférence en temps réel

#### PUBLIC CIBLE

- Analyste
- Statisticien
- Architecte
- Développeur
- Data scientist
- Machine Learning Engineer

#### PRE-REQUIS

- Connaissances générales sur le Machine Learning ainsi qu'en statistiques
- Notions de base en Python
- Avoir suivi la formation “Fondamentaux de la Data Science” (DSFDX)

#### METHODE PEDAGOGIQUE

Formation avec apports théoriques, échanges sur les contextes des participants et retours d'expérience pratique des formateurs, complétés de travaux pratiques et de mises en situation.

#### PROFIL DES INTERVENANTS

Cette formation est dispensée par un·e ou plusieurs consultant·es d'OCTO Technology ou de son réseau de partenaires, expert·es reconnus des sujets traités.

Le processus de sélection de nos formateurs et formatrices est exigeant et repose sur une évaluation rigoureuse leurs capacités techniques, de leur expérience professionnelle et de leurs compétences pédagogiques.

#### MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique.

Afin de valider les compétences acquises lors de la formation, un formulaire d'auto-positionnement est envoyé en amont et en aval de celle-ci.

En l'absence de réponse d'un ou plusieurs participants, un temps sera consacré en ouverture de session pour prendre connaissance du positionnement de chaque stagiaire sur les objectifs pédagogiques évalués.

Une évaluation à chaud est également effectuée en fin de session pour mesurer la satisfaction des stagiaires et un certificat de réalisation leur est adressé individuellement.

## **PROGRAMME PEDAGOGIQUE DETAILLE**

### **Jour 1**

#### **INTRODUCTION À LA COMPUTER VISION**

Les débuts de la CV, élément déclencheur

Frise chronologique de la CV et évolution du marché dans ce sens

Pourquoi le Deep Learning est-il devenu le standard ?

La révolution AlexNet

#### **LES DÉBUTS DE L'IMAGERIE DANS LE NUMÉRIQUE**

Rappel de traitement du signal

Représentation d'une image numérique

- Compression
- Encodage
- Résolution d'une image

Applications historiques

- Problème d'aliasing
- Balance des blancs
- Segmentation sémantique

Démonstration des transformations possibles d'une image selon les méthodes présentées

#### **MACHINE LEARNING CLASSIQUE APPLIQUÉ À L'IMAGERIE**

Représenter une image pour du Machine Learning

- Tableau de pixels
- Histogram de gradient et descripteurs SIFT (Scale-invariant feature transform)

Métriques d'évaluation

- Classification
- Détection d'objets
- Segmentation

Démonstration d'application du descripteur SIFT et mob pour implémenter la méthode

### **DEEP LEARNING APPLIQUÉ À L'IMAGERIE (PARTIE 1)**

Introduction aux CNN (Convolutional Neural Network)

Principe du transfert learning

Vision transformers et MaskFormers (HuggingFace)

Applications et métriques d'évaluation

- Classification : Théorie pour comprendre la classification dans le deep / Présentation des modèles à l'état de l'art
- Détection : Comprendre la détection d'objets dans une image / Présentation des modèles à l'état de l'art et particularités
- Segmentation (instance, semantic et panoptic) : Comprendre les différents types de segmentation d'objets (3 types) sur une image / Présentation des modèles à l'état de l'art et particularités

Démonstration d'application des différents types de segmentation et leur évaluation pour identifier leurs spécificités (ex: Bodypix)

Mise en pratique : "Implémenter, en mob, un modèle de Deep Learning effectuant de la détection et segmentation, en évaluant les performances du modèle - en passant par la méthode de transfer learning"

Mise en pratique : "Visualiser les filtres du modèle implémenté"

### **Jour 2**

### **DEEP LEARNING APPLIQUÉ À L'IMAGERIE (PARTIE 2)**

Présentation de Pytorch et Tensorflow

Applications à la vidéo

Mises en pratique :

- "Améliorer le modèle utilisé le jour 1 pour le même cas d'usage mais appliqué à une vidéo"
- "Évaluer la performance en temps de calcul de son modèle"

## **MODÈLES COMPARATIFS EN IMAGERIE**

### Contrastive learning (lien)

- Caractéristiques générales d'un ensemble de données sans label, en apprenant au modèle quels points de données sont similaires ou différents

### Zero-shot learning

- Lien entre la sémantique textuelle et ses représentations visuelles : Contrastive Language-Image Pretraining (CLIP) - OpenAI
- Guided Language to Image Diffusion for Generation and Editing (GLIDE)

### One-shot learning

- Facenet (Google, 2015): face recognition
- "Liveness detection" : distinction entre photo et personne physique (PyimagesSearch)
- RN siamois (lien)

### Few-shot learning

### Métriques d'évaluation

Démonstration : "Reconnaissance faciale en one-shot learning"

Mise en pratique : "Remplacer son modèle en appliquant un modèle CLIP (zero-shot-learning) et comparer les performances avec un modèle deep classique"

## **INTERPRÉTABILITÉ**

### Méthodes spécifiques aux images

### Shapley values & LIME

### Adversarial Attack

### Mises en pratique :

- "Vérifier que son modèle implémenté apprend bien ce que l'on souhaite qu'il apprenne"
- "Tromper le modèle qui doit détecter des personnes"

## **Jour 3**

## MODÈLES GÉNÉRATIFS

Data augmentation : Synthetic dataset

GAN

Modèles de diffusion

- DALL-E 2 : architecture - fonctionnement
- Stable diffusion : code avec Stability AI : site web , repo

Applications

- Unconditional image generation
- Inpainting : supprimer les objets indésirables d'une image ou recréer les régions manquantes d'images occluses (DFT) -lien paper
- Super-resolution (medium)
- Correction de défauts (ex: débruitage, augmentation des détails, amélioration de la qualité, niveau de lumière, flou...)

Evaluation des risques

- Applications frauduleuses: DeepFake
- Éthique: biais ethniques et sociaux

Démonstrations

- “Synthetic dataset, concrètement comment ça s'implémente”
- “Transfert de style avec du Deep Learning”

Mise en pratique : “Manipuler Stable Diffusion (prévoir des machines avec GPU)”

## IMAGERIE EMBARQUÉE

Contraintes

Optimisation de modèle (pour l'inférence)

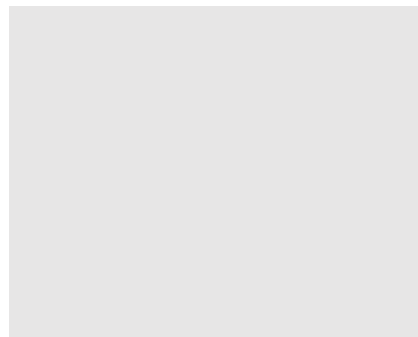
- Architecture de modèle spécifique (mobilenet, yolo) : Mobilenet, architecture du modèle et spécificité - avantage et inconvénient (rapide à inférer mais long à entraîner) / YOLO - YOLOv8, spécificité (ce qui le rend rapide et juste à la fois) - zoom sur temps d'entraînement et temps d'inférence
- Considérations énergétiques : Frugalité en ML / Knowledge distillation, quantization, weight pruning, morphnet
- Outils disponibles et accélérateurs d'inférence : TFlite / Tensorflow extended (TFX) / TensorFlow Lite Model Maker / CoralTPU / Tensor RT pour NVIDIA
- Mises en pratique : “Embarquer son modèle implémenté le jour 2 sur un Raspberry Pi ou sur une Coral” / “Effectuer de la détection en temps réel sur Raspberry Pi ou sur la Coral”

**BILAN ET CLÔTURE DE SESSION**

Revue des concepts clés présentés tout au long de la formation

Questions et réponses additionnelles

Evaluation et Clôture



---

**Accessibilité**

L'inclusion est sujet important pour OCTO Academy.

Nos référent·es sont à votre disposition pour faciliter l'adaptation de votre formation à vos besoins spécifiques.

Pour les contacter : [academy.accessibilite@octo.com](mailto:academy.accessibilite@octo.com)