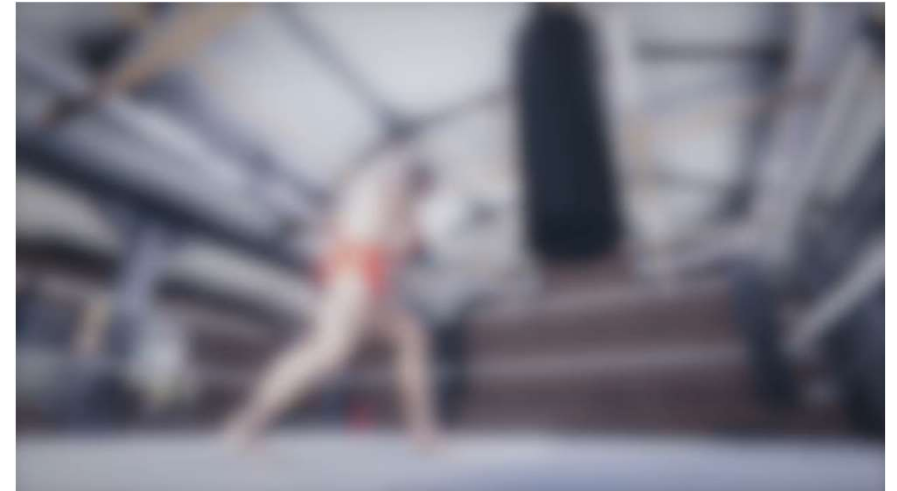


2007



2021

IA et Animation

ALEXANDRE MEYER

GDR IGRV

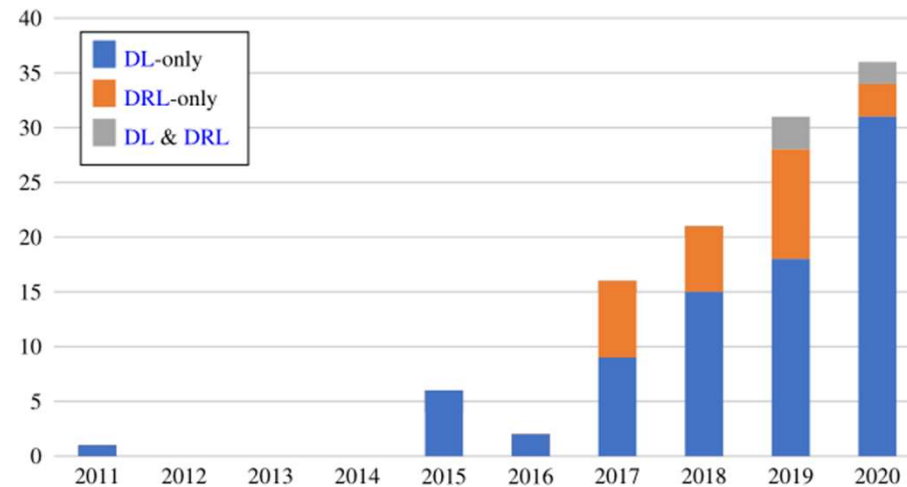
30/05/2023

SAARA / LIRIS



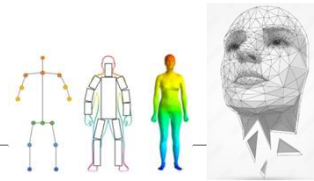
Apprentissage profond et animation

- Nous en sommes où ? Est-ce que c'était mieux avant ? 😊
- Est-ce qu'il y a des domaines de l'animation qui n'utilisent pas (encore) les réseaux ?
- Et ensuite ?



A Survey on Deep Learning for Skeleton-Based Human Animation
L. Mourot, L. Hoyet , F. Le Clerc, F. Schnitzler and P. Hellier, CGF2021

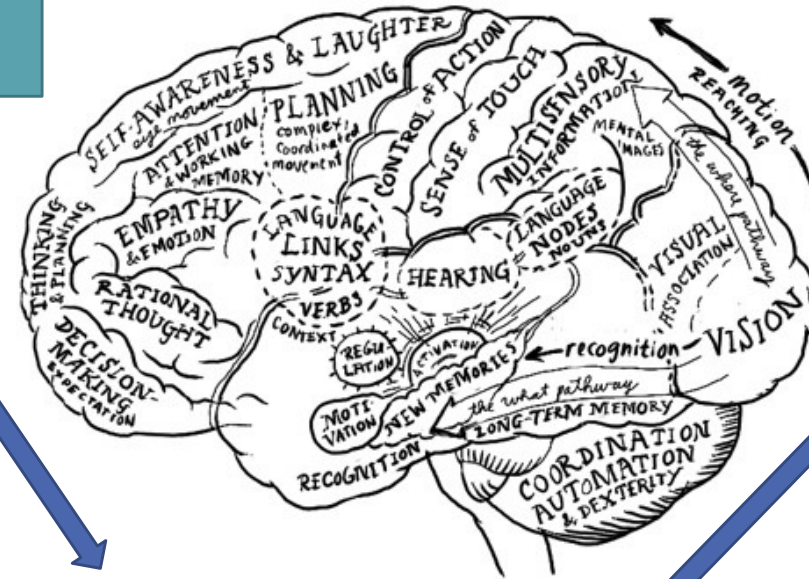
Introduction



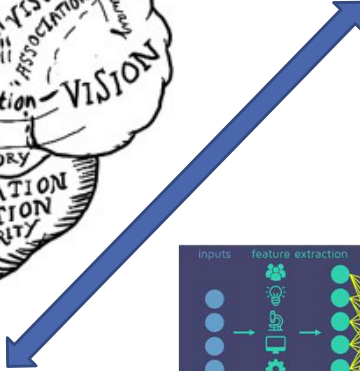
Vision par ordinateur



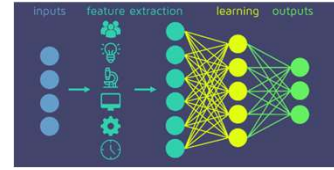
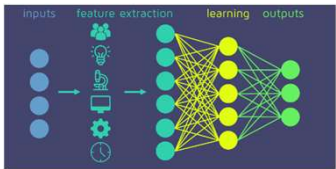
Angry Disgusted Fearful Happy



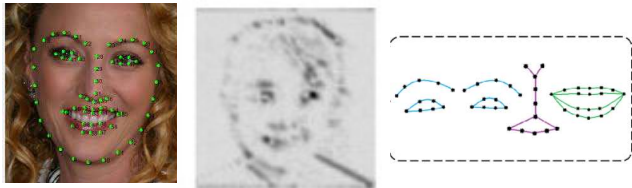
Graphique, Animation, Mouvement



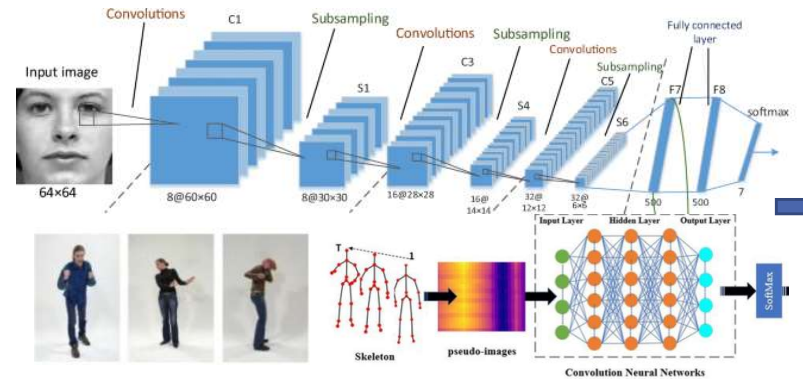
Intelligence artificielle (machine learning)



Vision par ordinateur



Squelette,
maillage,
action, etc.



Squelette,
maillage,
action, etc.



2010

2020

Modèles « compréhensibles »

- SVM
- Flot optique
- Croissance de région
- SIFT, SURF
- ...

Manque de capacité à généraliser

Apprentissage profond

- CNN ... Transformeur

→ Capture de mouvements monoculaire
Aide à la conduite, Génération d'images, ...

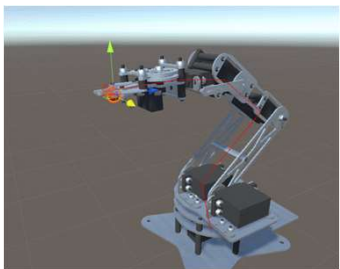
Animation de personnages virtuels



Conception humaine d'un modèle avec beaucoup de connaissances a priori

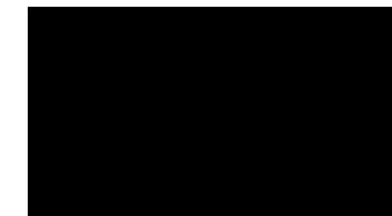
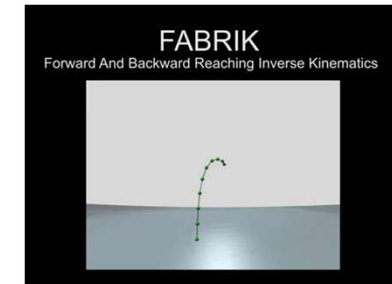


Animation procédurale

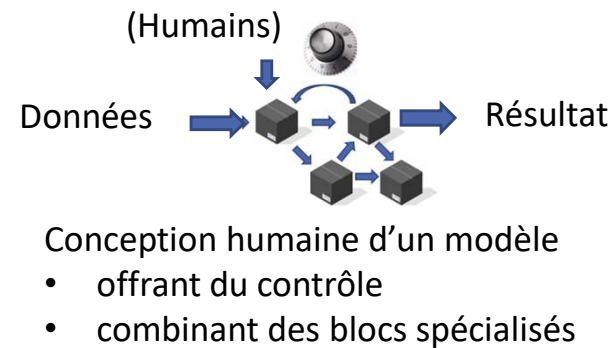
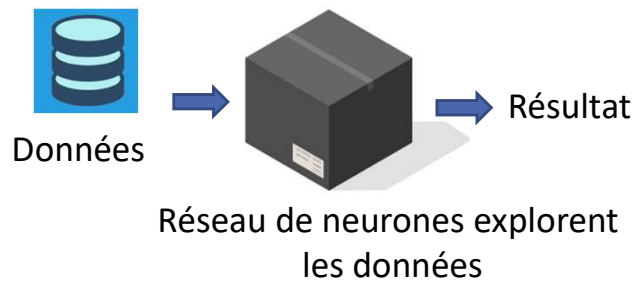


Basée sur la capture de mouvements

- Graphe d'animation, machine à états, motion matching
- Edition : cinématique inverse, ...
- Skinning
- ...



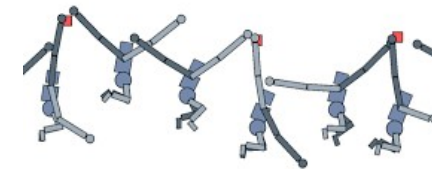
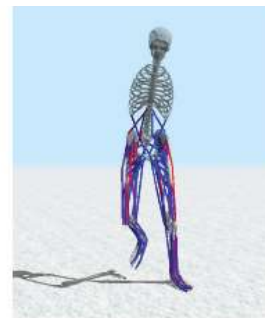
Animation de personnages virtuels

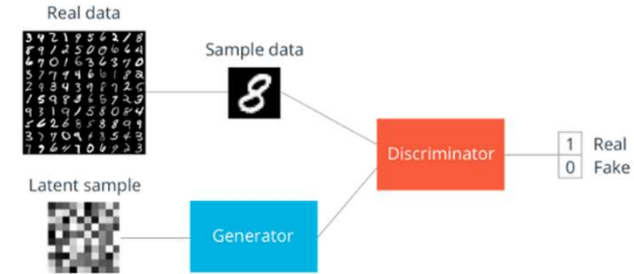
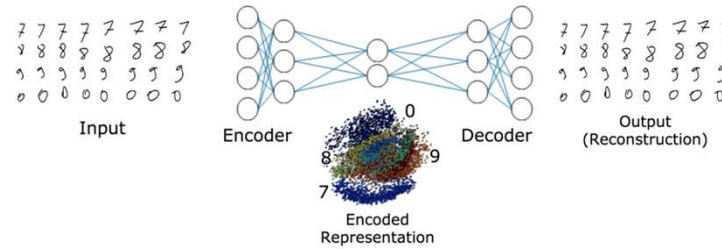
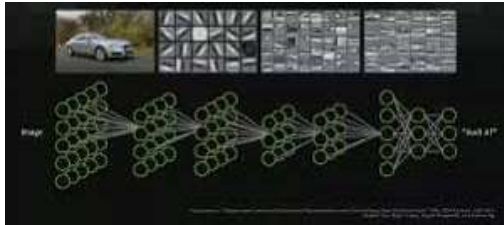


2015

Animation basée réseaux de neurones et animation apprise par renforcement dans un cadre physique

2020

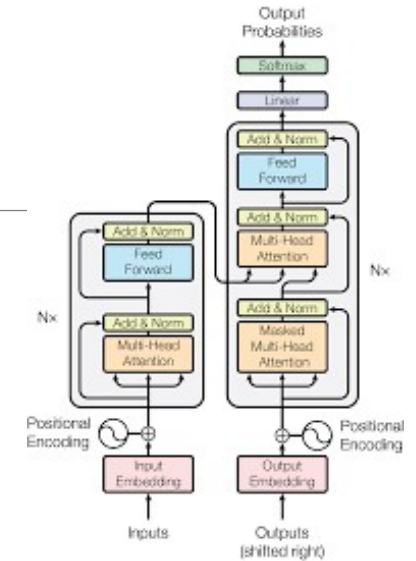




Types de réseaux

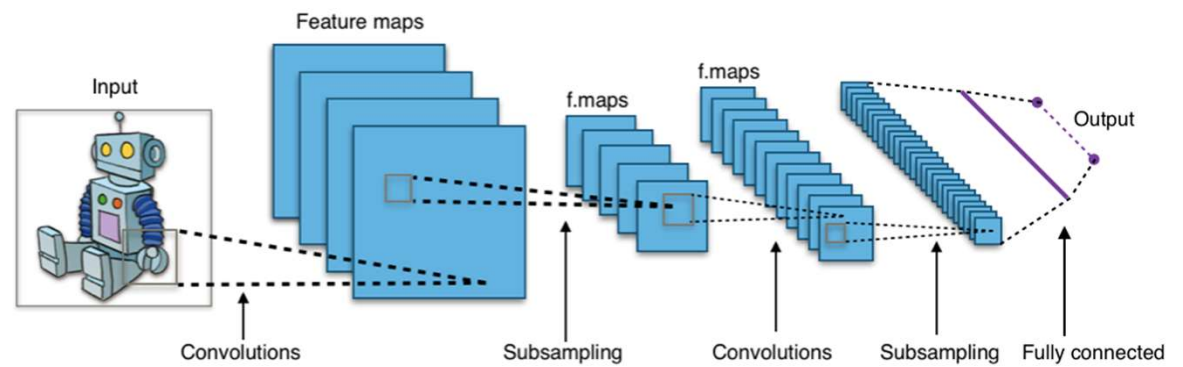
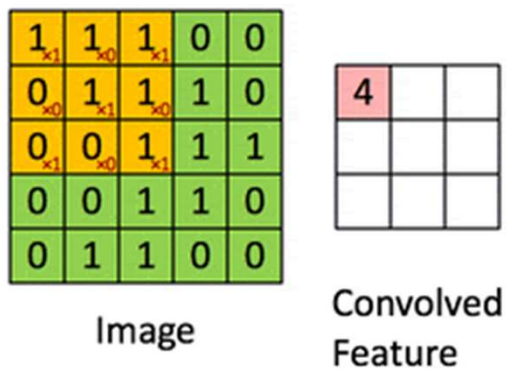
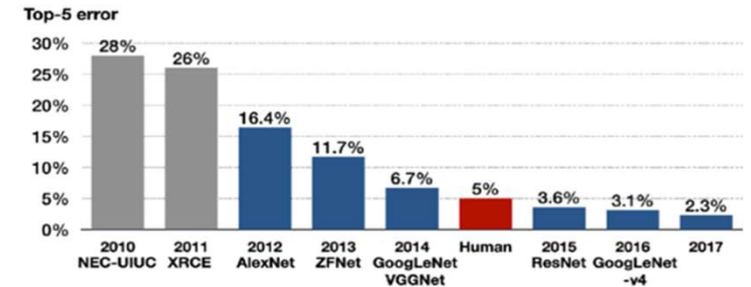
ET LEURS APPLICATIONS EN ANIMATION

- CNN
- AUTOENCODER
- ...
- TRANSFORMER



Réseaux convolutifs

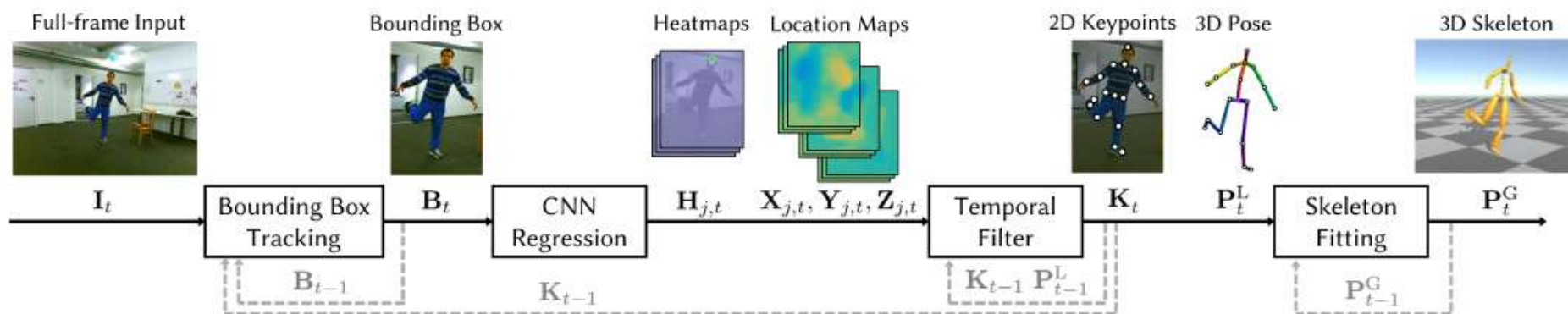
- Concours IMAGENET → mettre un label sur une image
 - 10 millions d'images avec 11 000 classes
 - Réseaux de convolution ou CNN, ConvNet
- VGG16 : 160 millions de paramètres, 16 séries de convolution



Vnect et Xnect

Capture temps réel de la pose à partir d'une unique camera RGB

- Tracking de la boîte englobante
- Heatmaps des articulations
- Location map pour chaque articulations $X_j Y_j Z_j$
- Articulations 2D \rightarrow posture 3D



Squelette 3D : optimisation ou non

- (Vnect) Optimisation par moindres carrés

- Aspect temporel
- Garantir les longueurs
- Lissage temporel
- ➔ Temps de calcul « long »

- (Xnect) Tout le processus dans un réseau

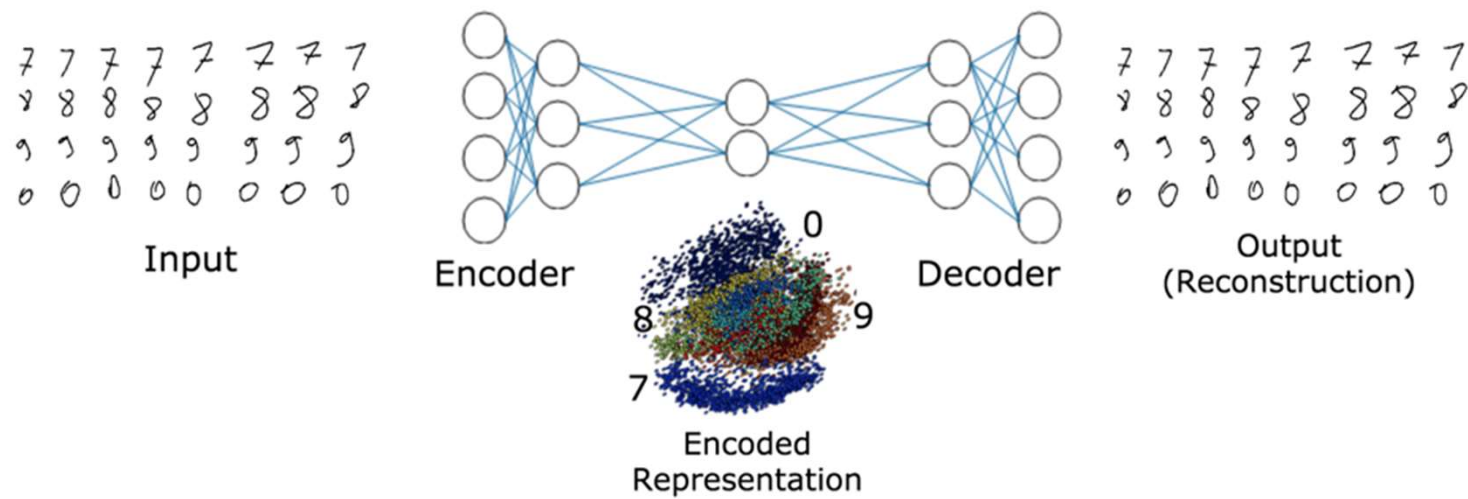
- ➔ plus rapide à calculer
- Plus robuste (multi-personnes)



XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera

Dushyant Mehta¹, Oleksandr Sotnychenko¹, Franziska Mueller¹, Weipeng Xu¹, Mohamed Elgharib¹, Pascal Fua², Hans-Peter Seidel¹, Helge Rhodin^{2,3}, Gerard Pons-Moll¹, Christian Theobalt¹



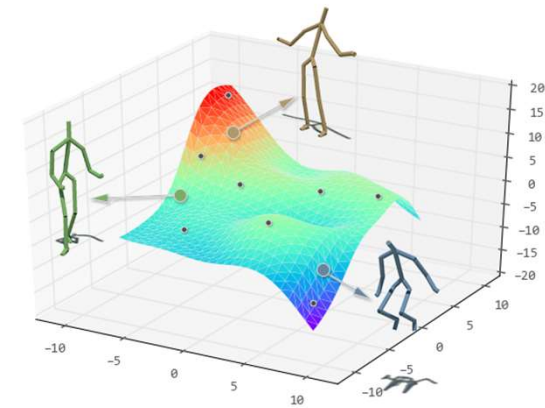
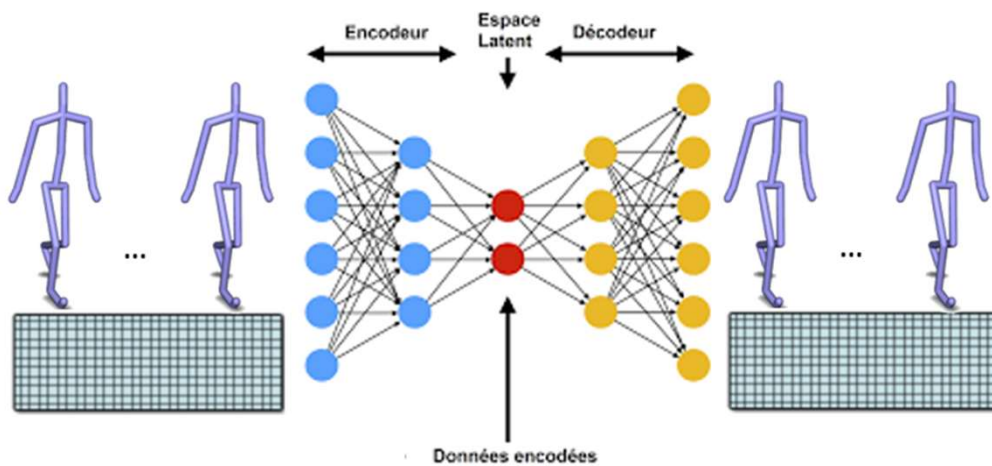


AutoEncoder

ESPACE LATENT

AutoEncoder

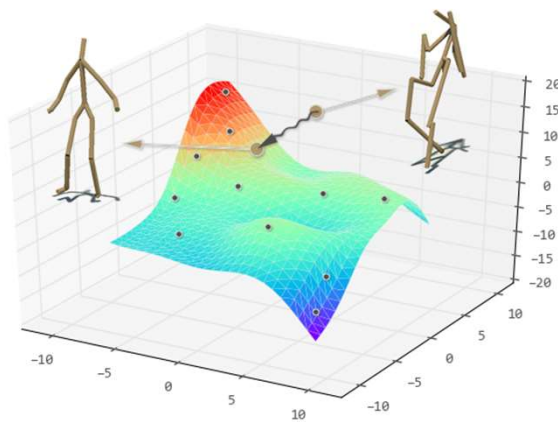
- Entraîne un réseau à ressortir les mêmes données qu'en entrées
 - Encodeur + Décodeur
 - Construire un espace latent compact, efficace
 - Variational AutoEncoder (VAE) : contraint le code sur une distribution



Chaque point représente une animation complète

Projection : correction d'erreurs

- Projection sur l'espace caché corrige les problèmes
 - Marqueurs manquant
 - Kinect
 - Step motion

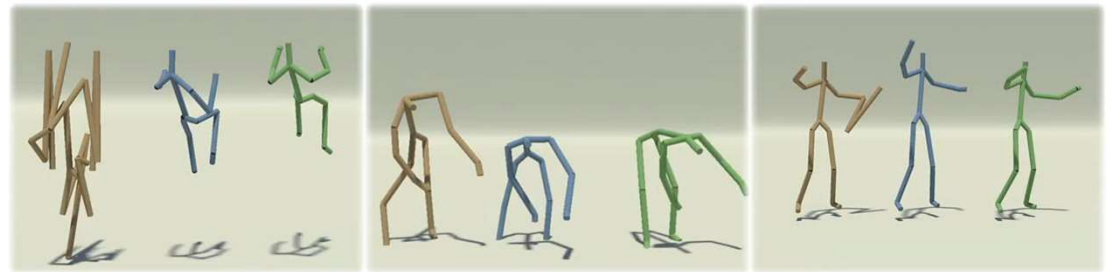


Learning Motion Manifolds with Convolutional Autoencoders



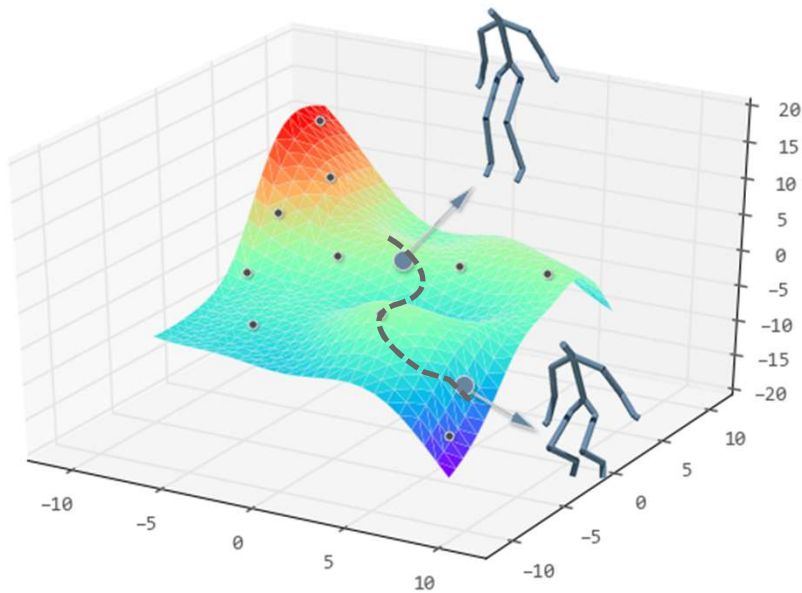
Daniel Holden, Jun Saito, Taku Komura, Thomas Joyce

Edinburgh University, Marza Animation Planet



Interpolation

- Interpolation dans l'espace latent évite les animations invalides

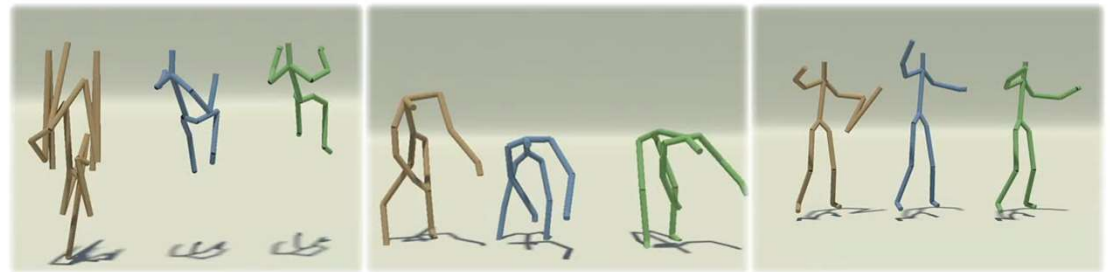


Learning Motion Manifolds with Convolutional Autoencoders



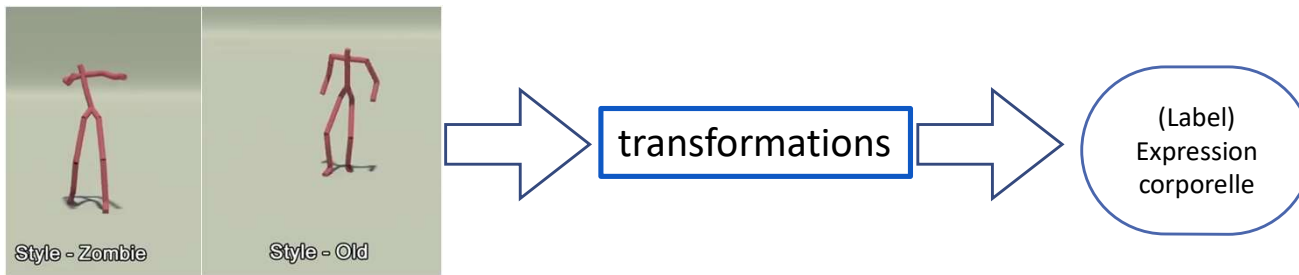
Daniel Holden, Jun Saito, Taku Komura, Thomas Joyce

Edinburgh University, Marza Animation Planet

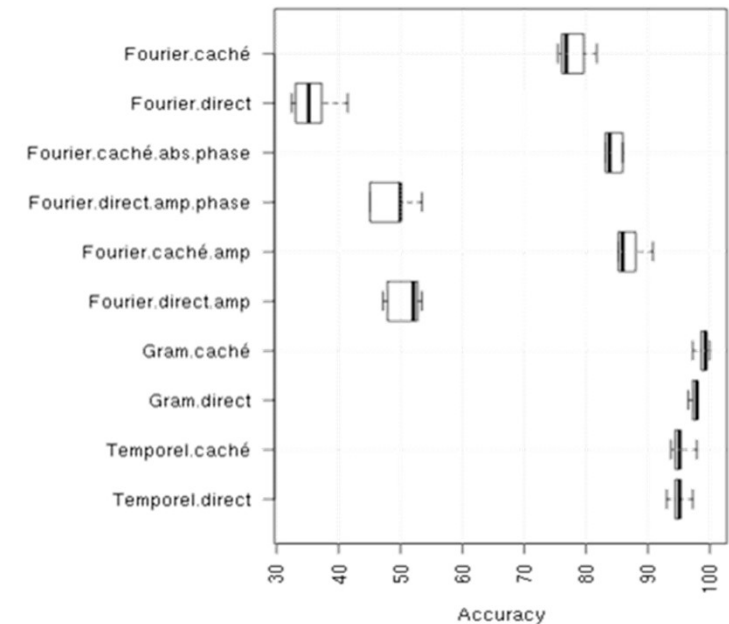
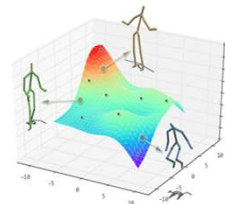


Reconnaissance d'expressions

- Espace latent : représentation efficace ?
 - Un algo de reconnaissance devrait le confirmer
 - Test de plusieurs transformations



- Bilan
 - Articulations représentées avec des positions relatives
 - Matrice de Gram (« exhausteur » de donnée)
 - **Autoencodeur améliore à chaque fois les résultats**



Premier bilan

Il y a des choses à faire avec les espaces latents !

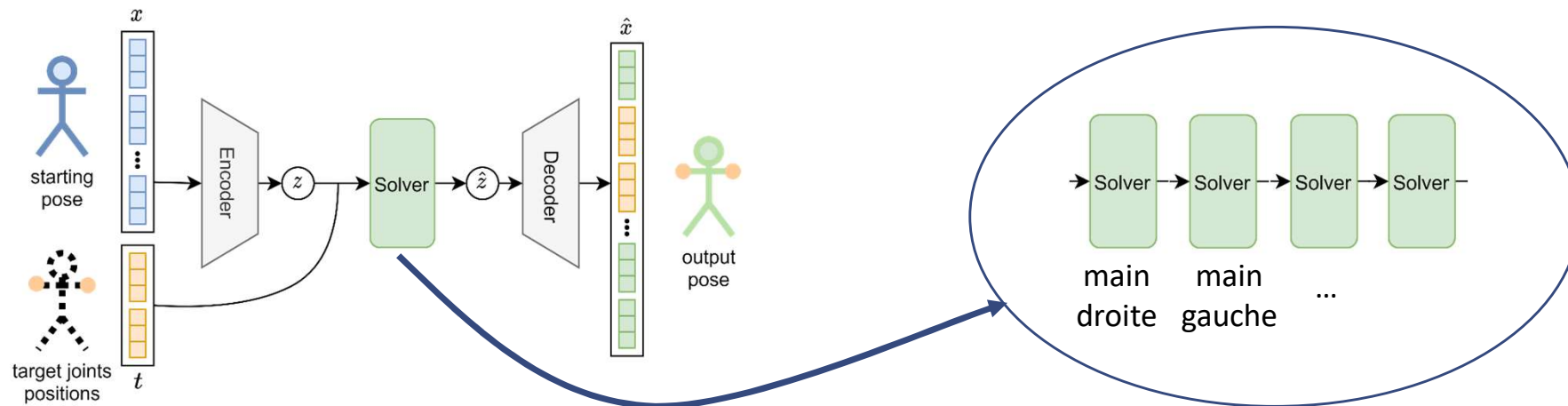
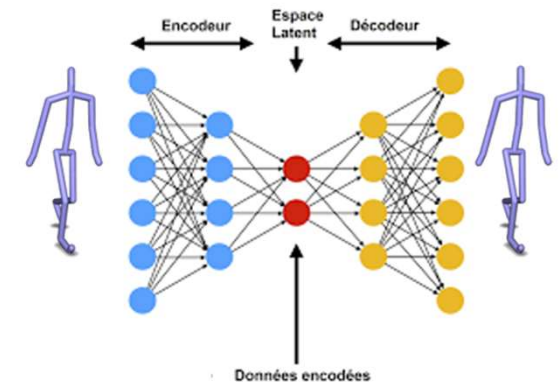
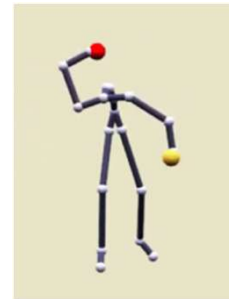
- Dataset : quelques heures d'animations suffisent (CMU)
- Temps d'entraînement : quelques heures pour un AE d'animation

Mais quelle représentation des animations en entrée ?

- Position 3D : longueur non garantie
- Angles vs quaternions ? → QuaterNET
- Tous les squelettes doivent être identiques
- Comment savoir si l'espace latent contient les bonnes propriétés ?
TSNE, Tâche de reconnaissance donne une indication
- Comment « utiliser » un espace latent pour modifier des paramètres ?
 - ...

Edition de poses dans l'espace latent

- Cinématique inverse
 - Outils important pour l'édition
- Autoencodeur de poses
 - Représentation latente de l'espace des poses
 - Se « déplacer » dedans pour générer des poses satisfaisant des choix d'un utilisateur



Edition de poses dans l'espace latent

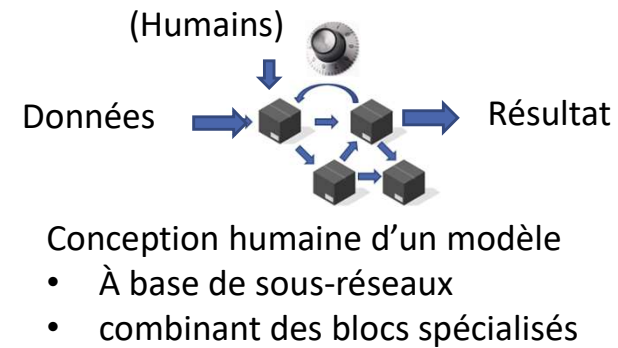
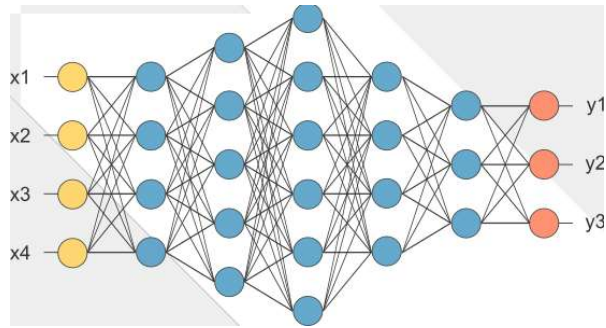
Learning-based pose edition for efficient and interactive pose design

Léon Victor*, Alexandre Meyer**, Saïda Bouakaz**

Univ Lyon, LIRIS, UMR CNRS 5205

** Université Claude Bernard Lyon 1

* INSA Lyon



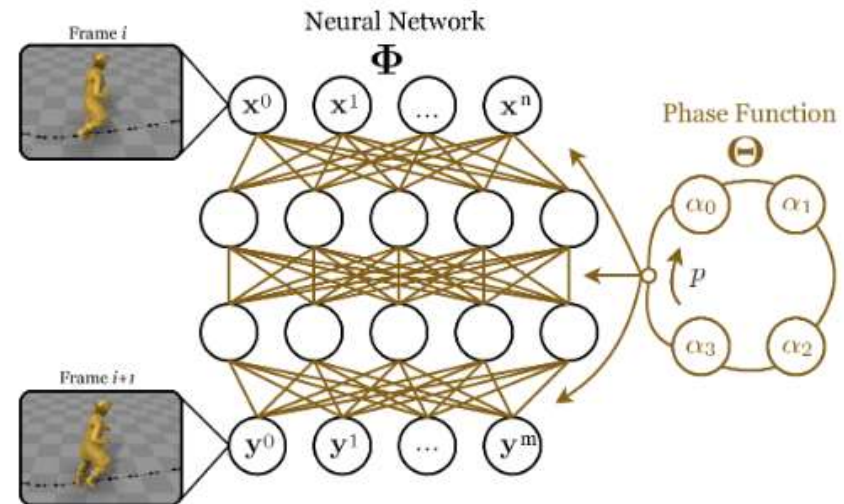
Réseaux experts

PLUSIEURS RÉSEAUX SPÉCIALISÉS EN INTERACTION

Réseaux experts : Phase-Funct. NN

Contrôleur de déplacement d'un personnage dans un environnement virtuel

- Entrée
 - Trajectoire passée du personnage
 - Pose actuelle (+ vitesse)
 - Trajectoire désirée
- Réseaux avec Fonction de Phase
- Sorties
 - Pose suivante



- La bonne idée ici : poids dépendent de la phase de la marche

Réseaux experts : Phase-Funct. NN

Contrôleur de déplacement d'un personnage dans un environnement virtuel

- Entrée
 - Trajectoire passée du personnage
 - Pose actuelle (+vitesse)
 - **Trajectoire désirée au sol**
- Réseaux avec Fonction de Phase
- Sorties
 - Pose suivante

Phase-Functioned Neural Networks for Character Control

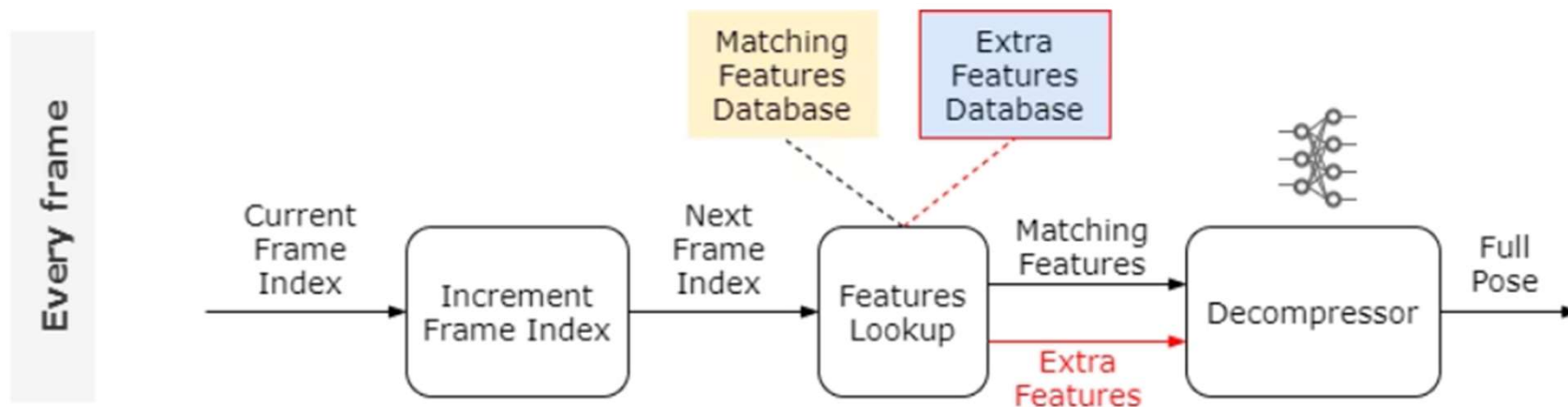
Daniel Holden¹, Taku Komura¹, Jun Saito²

¹University of Edinburgh, ²Method Studios

- La bonne idée : poids dépendent de la phase de la marche
 - Plus simple qu'un Conditional GAN

Réseaux experts : Learned Motion Matching

- Phase-F. NN produit toute la pose avec un unique réseau → manque de détails
- La sélection par recherche de mouvement propose plusieurs réseaux
 - Réseau1 → les caractéristiques de base de la pose suivante : pieds, pelvis et vitesse de chacun
 - Réseau2 → « extra features » : code d'un auto-encodeur
 - Décodeur → la pose

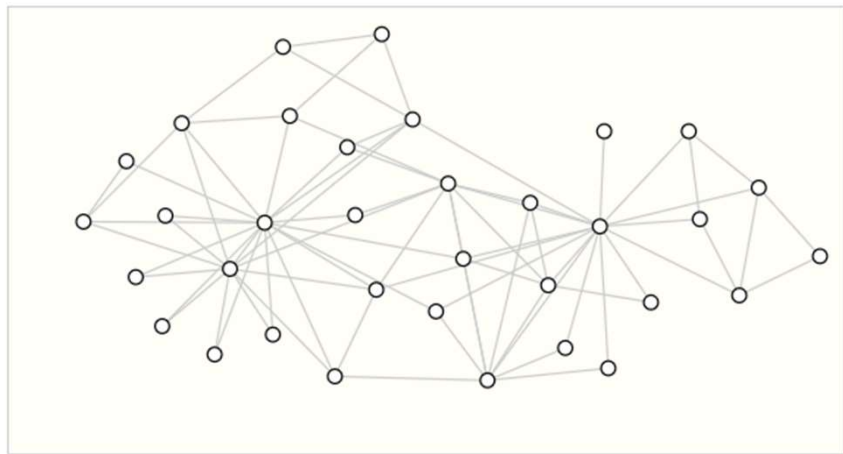
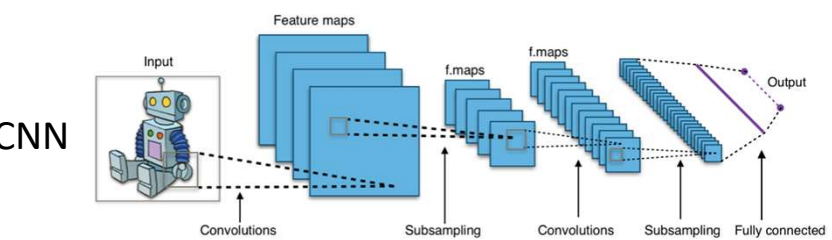
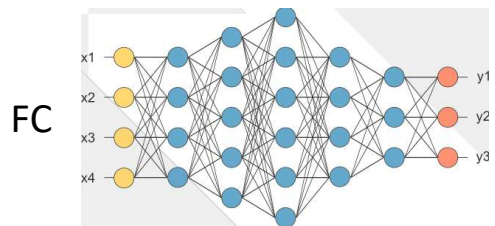


Réseaux experts : Learned Motion Matching

	Sélection par recherche de mouvement	Décompresseur	Sélection apprise de mouvement
Animations sources	461.0 Mo	-	-
Caractéristiques de sélection	129.8 Mo	129.8 Mo	-
Caractéristiques complémentaires	-	47.2 Mo	-
Paramètres du Décompresseur	-	2.4 Mo	2.4 Mo
Paramètres du Séquenceur	-	-	1.4 Mo
Paramètres du Projecteur	-	-	13.0 Mo
Total	590.8 Mo	179.4 Mo	16.8 Mo

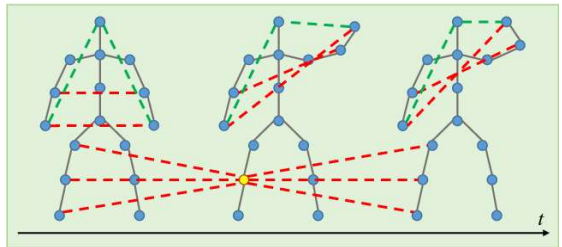


Learned Motion Matching
Holden et al. SIGGRAPH 2018



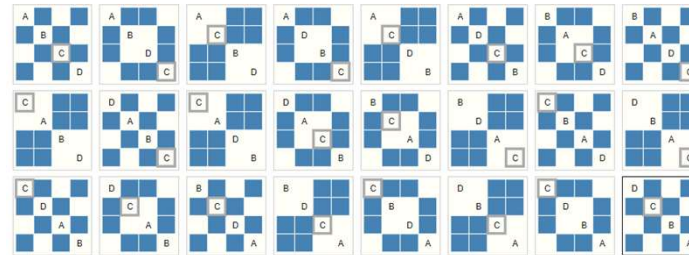
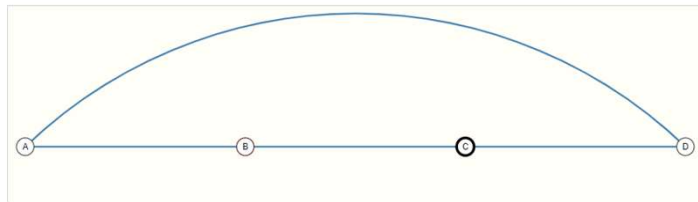
Graph Neural Net (GNN)

DE NOMBREUX PROBLÈMES SE MODÉLISENT AVEC UN GRAPHE

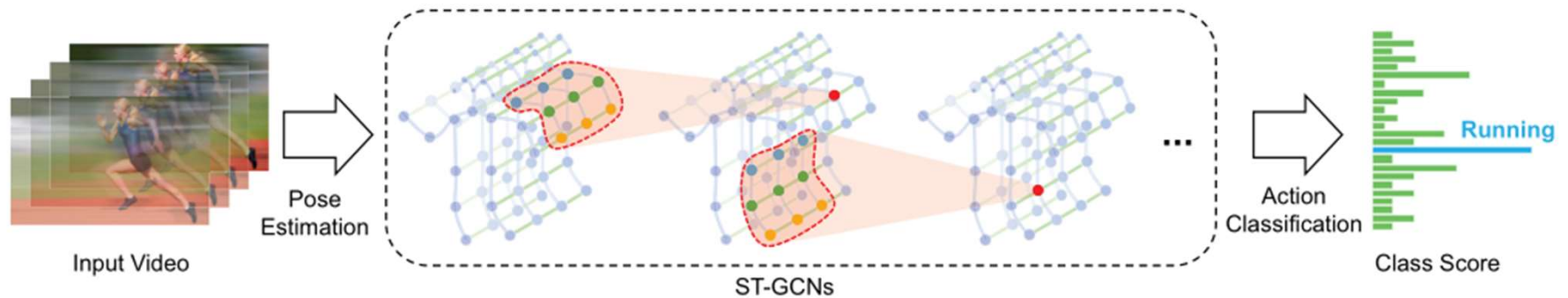


GraphNN

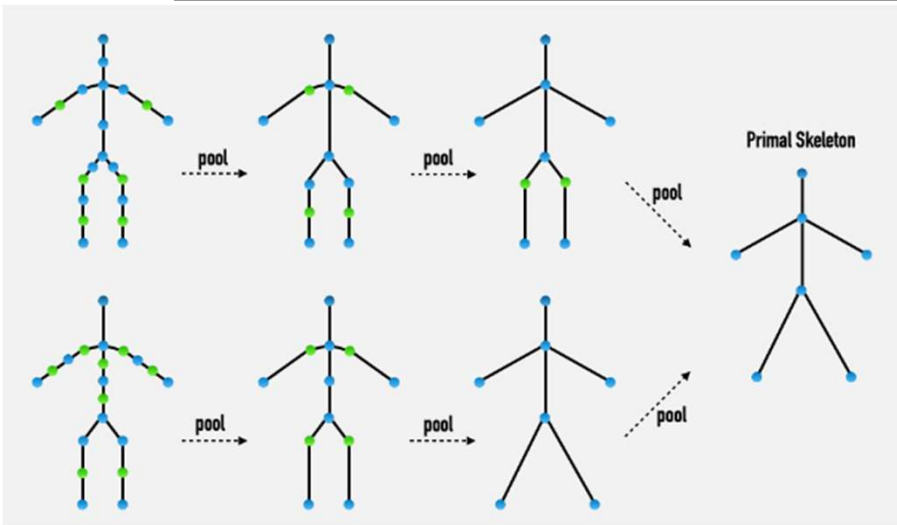
- Difficulté avec les graphes : permutation de la matrice d'adjacence → encodeur



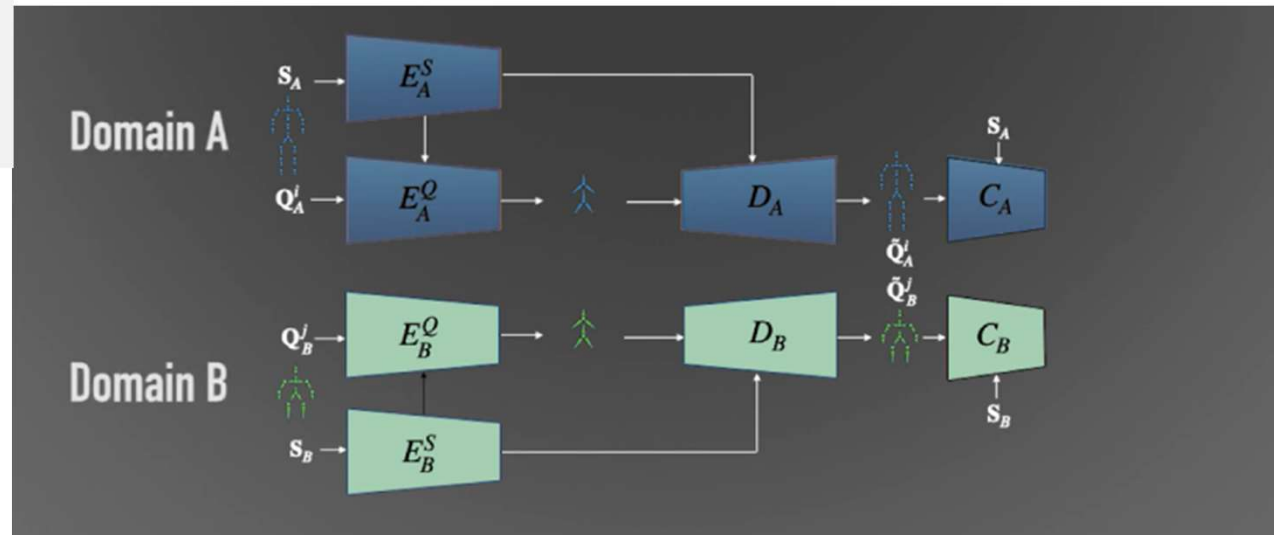
- Graphe de convolution spatiale et temporel (ST-GCN)



Retargeting : unifier les squelettes



- Autoencoder
 - Static : squelette au repos
 - Animation
- Encodeur et décodeur sont des GCN



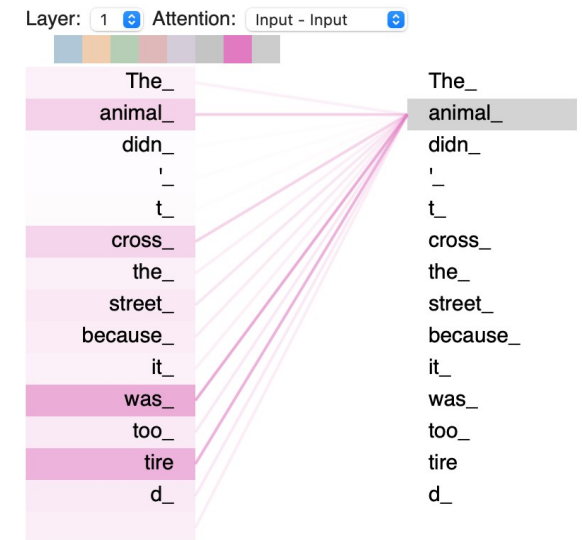
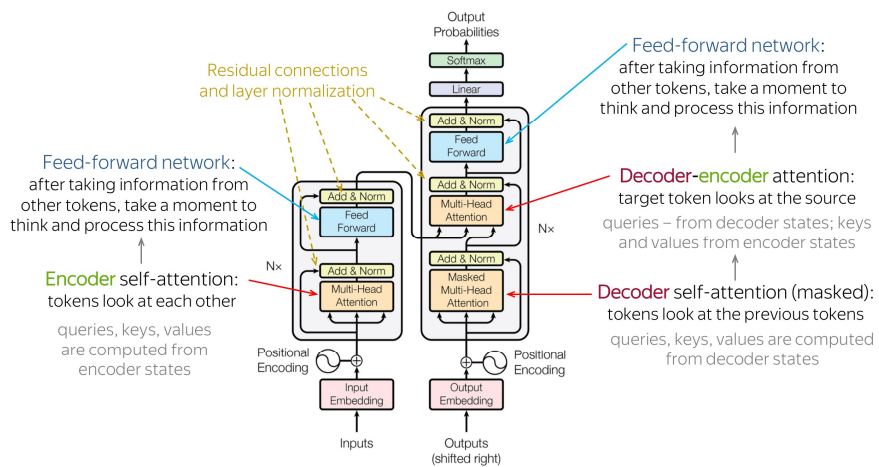
Retargeting : unifier les squelettes

Skeleton-Aware Networks for Deep Motion Retargeting

Kfir Aberman*, Peizhuo Li*, Dani Lischinski, Olga Sorkine-Hornung,
Daniel Cohen-Or, Baoquan Chen

(The video contains voice)





Transformer

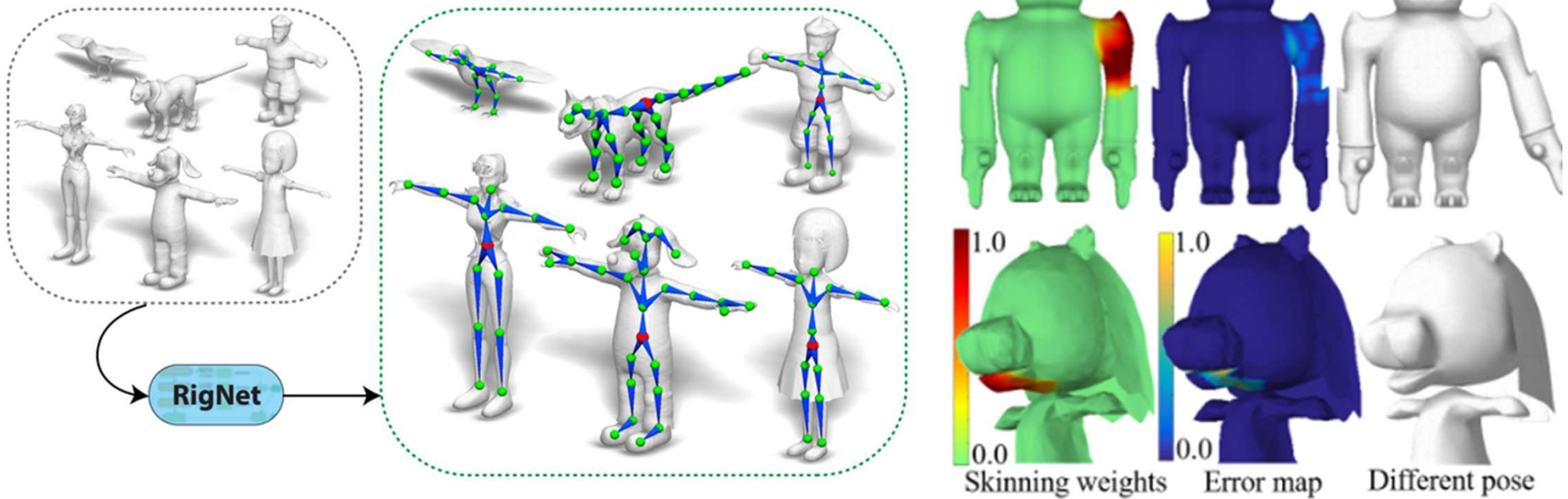
ET SURTOUT LES MODULES D'ATTENTIONS



Le réseau qui monte ...

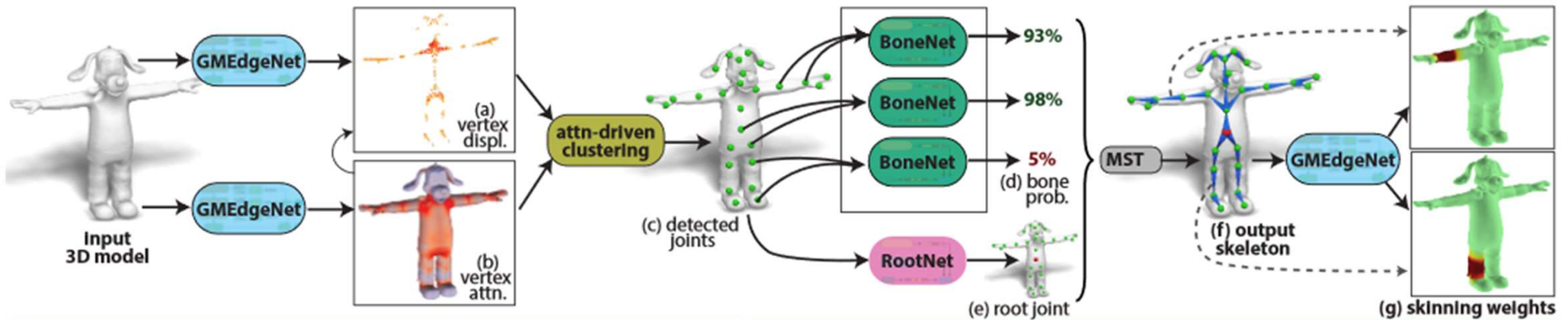
RIGS : squelette et poids du skinning

- Entrée: maillage 3D d'un personnage
- Sortie: squelette et les poids du skinning



RIGS : squelette et poids du skinning

- Maillage et squelette : GraphNN
- Modèle d'attention
 - prédit le lien entre chaque os du squelette et chaque sommet du maillage



Skeletal **joint** prediction

Skeleton **connectivity** prediction

Skinning prediction

RIGS : squelette et poids du skinning



Génération de mouvements

- “Motion In-betweening via Two-stage Transformers”
Autoencoder avec Transformers

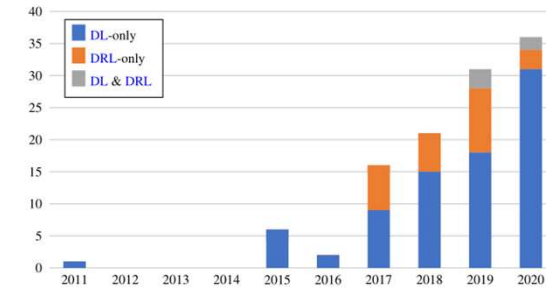
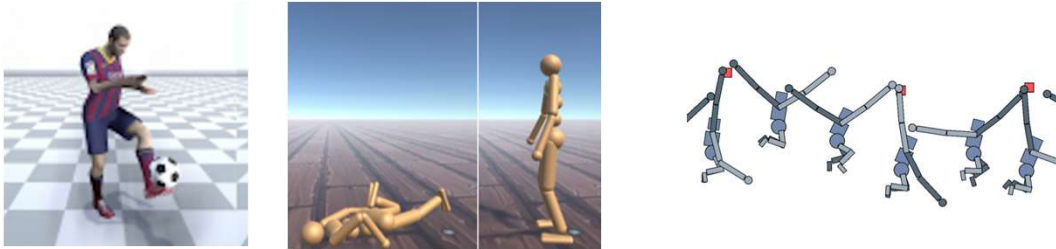
Motion Inbetweening via Two-stage Transformers

Jia Qin^{1,2} Youyi Zheng¹ Kun Zhou¹

¹State Key Lab of CAD&CG, Zhejiang University

²Netease Fuxi AI Lab

SIGGRAPH 2022



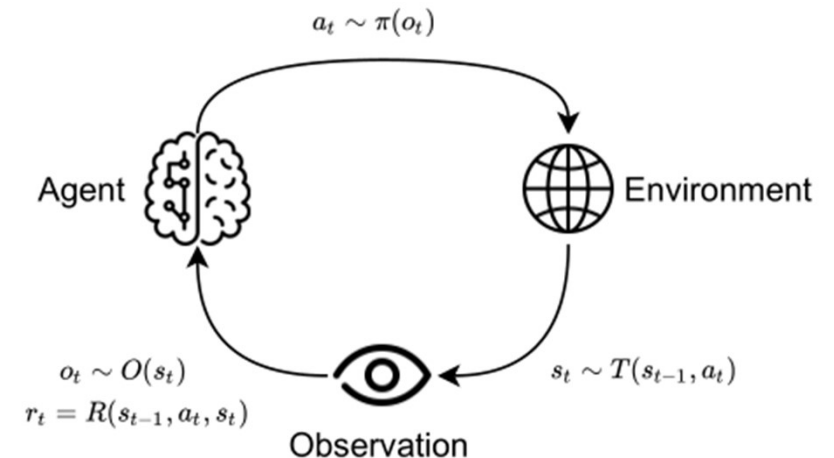
Apprentissage par renforcement (RL)

PROBLÈME DES DONNÉES → RL EXPLORE

RL : un exemple en animation

Environnement avec une simulation physique d'un ragdoll (personnage avec des solides rigides)

- Agent doit apprendre à déplacer le ragdoll
 - Agent réalise des actions
donne des forces / moment angulaire aux articulations
 - Agent observe l'environnement
= position + rotation de toutes les membres du corps
 - Environnement note (récompense)
position plausible, énergie, distance parcourue
- Algorithme de RL se basent sur les expériences passées
→ réseaux !!!
 - Réseau pour prédire la récompense en fonction de l'action
 - Réseau pour prédire la bonne action
 - PPO, SAC, etc.
 - Curriculum-driven : exercices de plus en plus difficiles



RL : from scratch

- DeepMind



RL : personnage / simulation / réalisme

- CALM : apprentissage à partir d'une séquence de capture de mouvement sans label
 - Un encodeur de mouvement + une politique de bas niveau (décodeur) entraînés par RL
 - ➔ convertir une séquence de capture de mouvement en actions contrôlant le personnage simulé
 - Une politique de haut niveau entraîné par RL
 - ➔ permettre le contrôle de la direction
 - Combinaison des étapes 1 et 2 à l'aide d'une simple machine à états finis

In video games, virtual characters react to user input and produce diverse motions

Deuxième bilan ...

- Revisite de tous (?) les anciens problèmes avec de l'apprentissage
 - Avec des améliorations
 - Mais aussi des temps d'apprentissage plus long puisque le réseau doit souvent tout réinventer

→ besoin de recycler des modèles déjà entraînés
- Techniques
 - GAN ? Réseaux de diffusion ?
- Et maintenant ?



Paper



Code



Data

Texte + animations

- AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars
Hong et al. SIGGRAPH 2022
<https://github.com/hongfz16/AvatarCLIP>



A *tall and skinny female soldier* that is *arguing*.



A *skinny ninja* that is *raising both arms*.

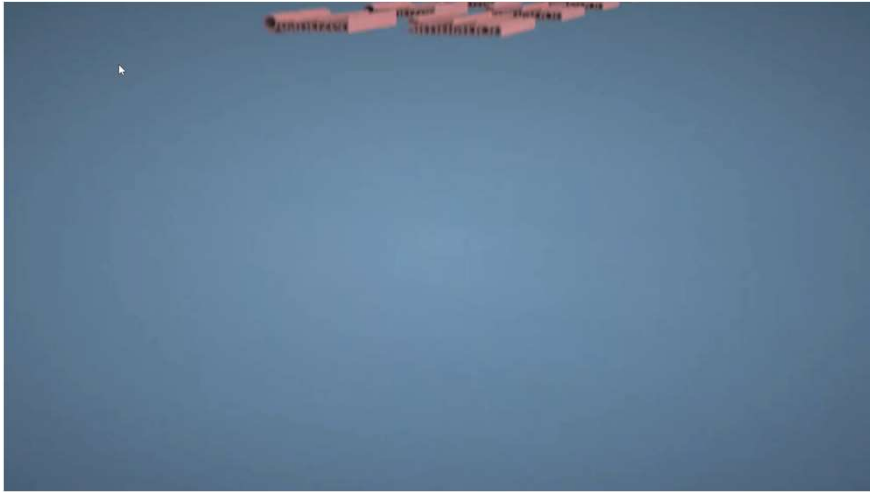


An *overweight sumo wrestler* that is *sitting*.



A *very skinny general* that is *eating hamburger*.

- Write-An-Animation: High-level Text-based Animation Editing with Character-Scene Interaction
Zhang et al . Eurographics 2021



Moteurs physiques rapides (GPU) et différentiables existent !

- Par exemple, 1 milliard de particules (80Go) avec Taichi Element
- ➔ **QuanTaichi: A Compiler for Quantized Simulations (SIGGRAPH 2021)**

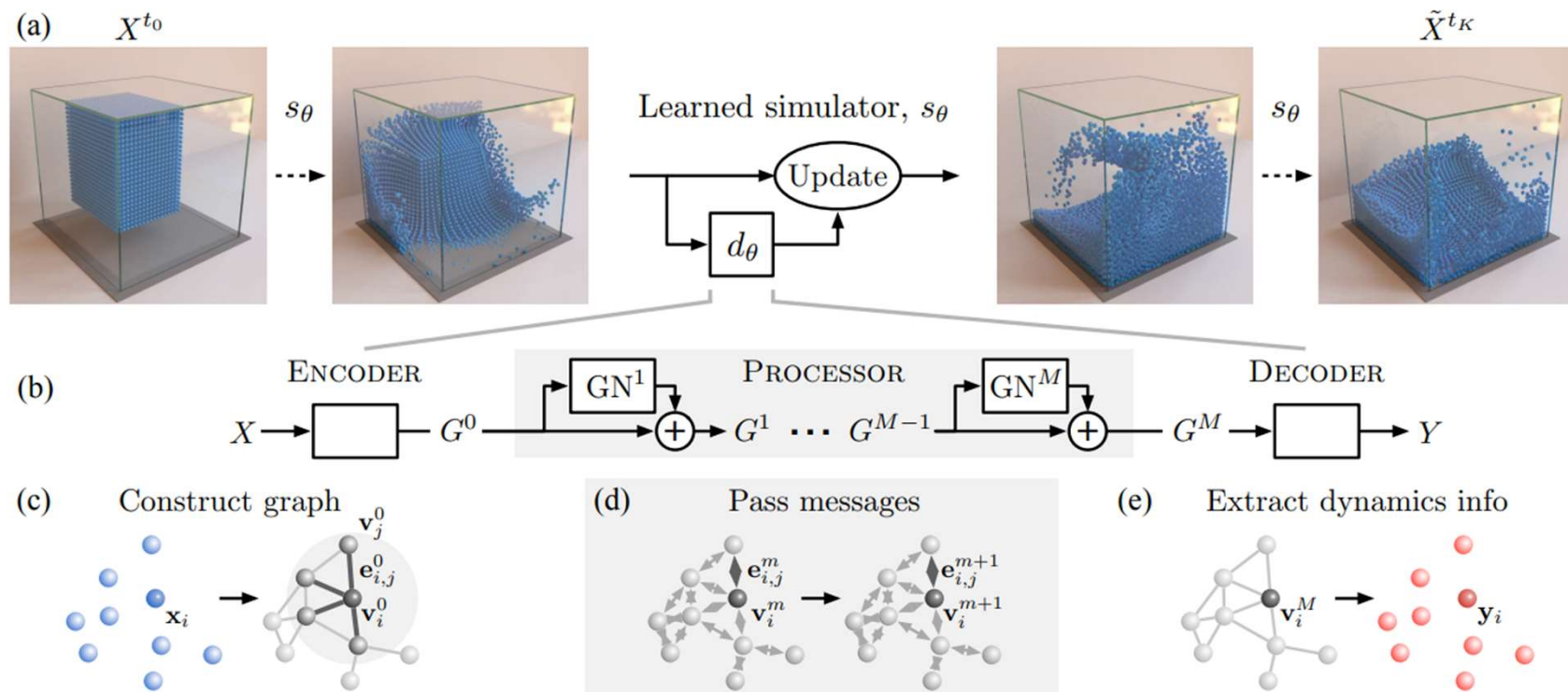
Simulation physique

DES OUTILS POUR DE LA SIMULATION INTENSIVE

SIMULATION + LEARNING ➔ LE NOUVEL ELDORADO ?

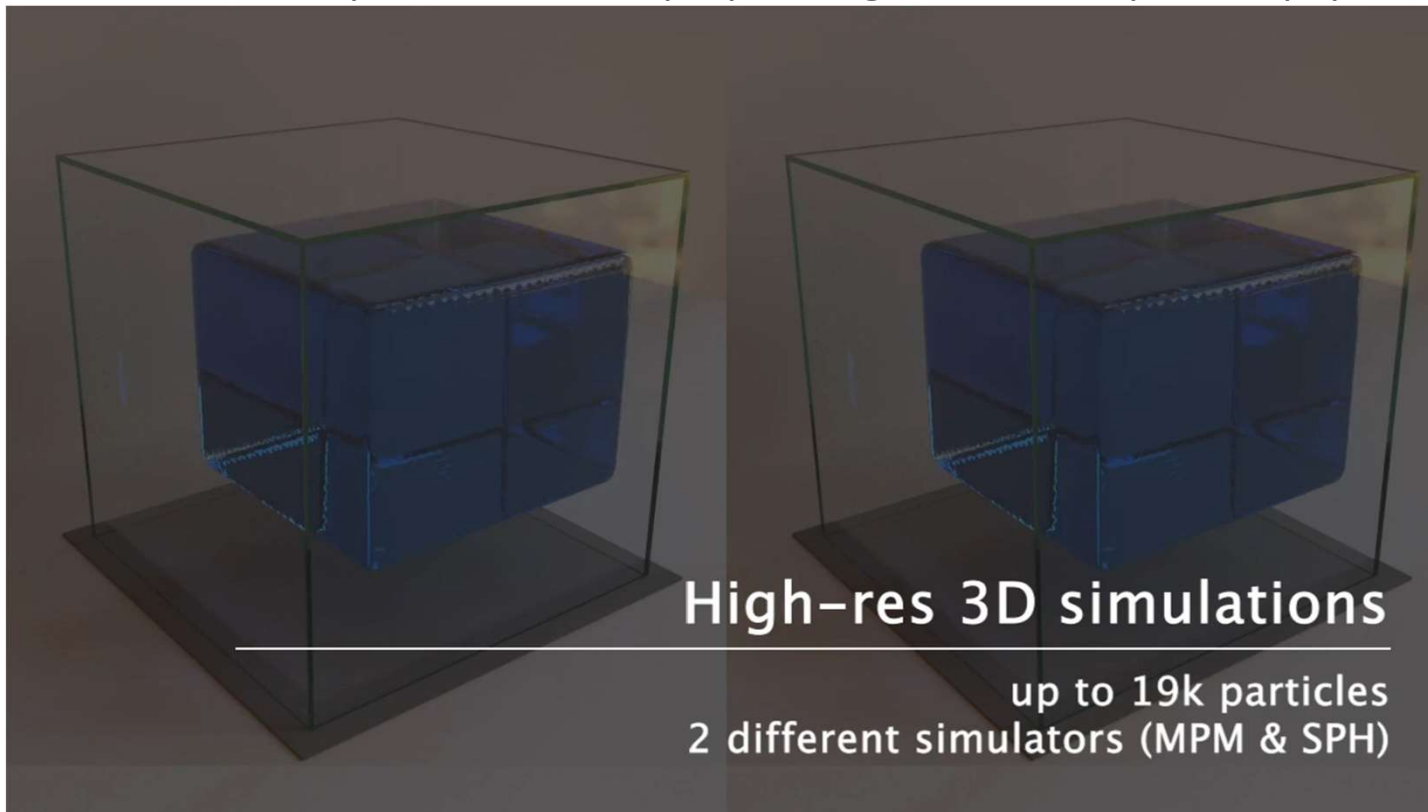
Simulation physique

- GNN / Encoder-decoder → gagner du temps de calcul



Simulation physique

- GNN / Encoder-decoder → preuve de concept (pas de gain de calcul pour ce papier)



Simulation physique de tissus

Subspace Neural Physics: Fast Data-Driven Interactive Simulation

Daniel Holden

UBISOFT
LA FORGE

Bang Chi Duong

UBISOFT
LA FORGE

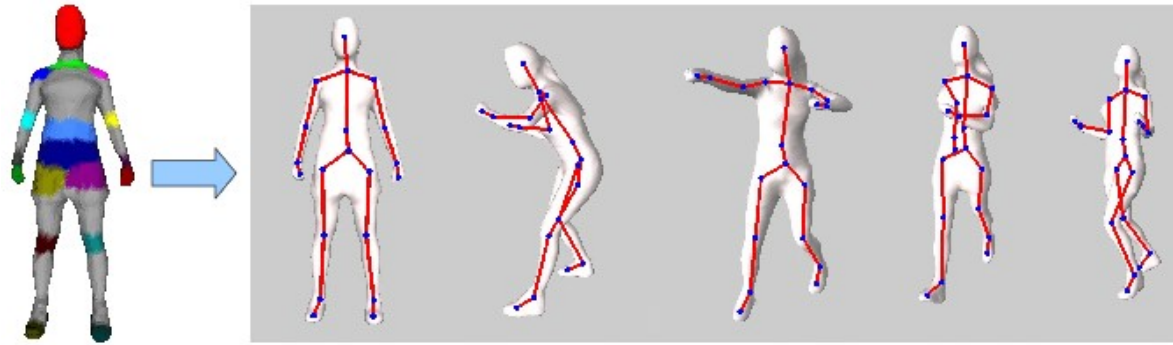
Entre x300 et x5000
par rapport à la simulation

Sayantana Datta

 McGill

Derek Nowrouzezahrai

 McGill

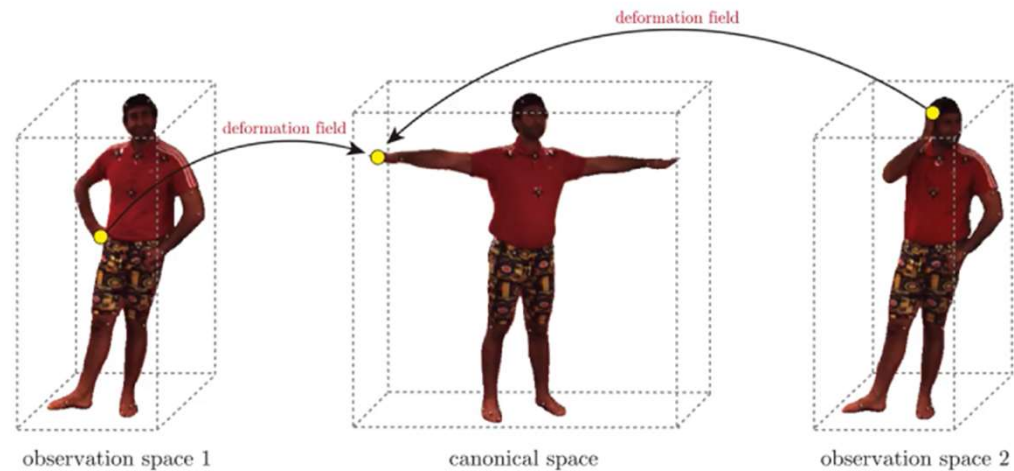


Apparence ... NERF

REPLACEMENT MAILLAGE/TEXTURE PAR NERF

FUSION SQUELETTE + APPARENCE

Nerf and Co.



- Animatable Neural Radiance Fields for Modeling Dynamic Human Bodies
Peng et al. ICCV 2021
- HumanRF: High-Fidelity Neural Radiance Fields for Humans in Motion
Işık et al. SIGGRAPH 2023
- NeRFahedron: A Primitive for Animatable Neural Rendering with Interactive Speed
Sin et al. SIGGRAPH 2023. <https://zackarysin.github.io/NeRFahedron>
- ...

Nerf Animé

Animatable Neural Radiance Fields for Human Body Modeling

Sida Peng, Junting Dong, Qianqian Wang, Shangzhan Zhang

Qing Shuai, Hujun Bao, Xiaowei Zhou



浙江大學
ZHEJIANG UNIVERSITY



Cornell University

Nerf Animé



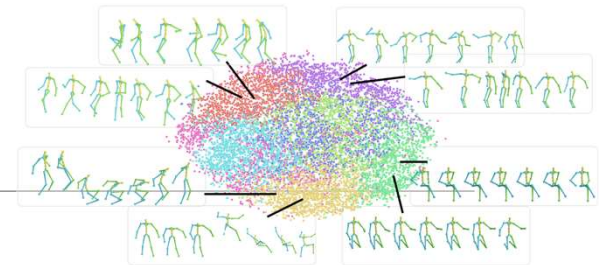
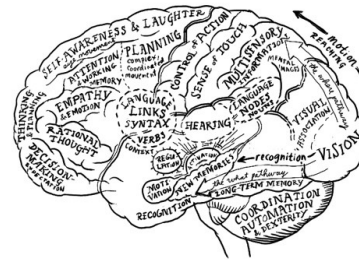
HumanRF

High-Fidelity
Neural Radiance Fields
for Humans in Motion

Mustafa Işık, Martin Rünz, Markos Georgopoulos, Taras Khakhulin,
Jonathan Starck, Lourdes Agapito, Matthias Nießner

 synthesia  UCL   ACMSIGGRAPH

Et maintenant ?



- Grands modèles de langage ... et pourquoi pas un grand modèle d'animations ? animGPT ?
 - Les données : ... youtube, les films ...
 - Interaction avec des objets / les mains ? (pb de vision)
 - Animations collaboratives entre 2 ou 3 personnes ?
- Mais comment avoir un contrôle sur l'espace latent ?
 - Multimodalité ? Texte + animation : est-ce indispensable ?
- Convergence squelette + apparence ?
 - Capture de l'apparence devient possible ? Mais qu'en disent les animateurs ?
- Animation physique
 - Accélérer l'exploitation des calculs pour des applications temps réel
 - Et si toutes les primitives de rendu deviennent des NERF ? Collision ? Déformation ?

« Une course dynamique, suivi d'un saut en longueur élégant et aérien avec un record, puis un geste rageur de victoire »