

Playable Cinema

—

Projet RCDAV HES-SO, septembre 2023

Requérant : Douglas Edric Stanley, HEAD – Genève

Durée : janvier à décembre 2024 (12 mois)

—

Résumé

Ce projet propose d'utiliser des techniques d'intelligence artificielle émergentes pour développer une méthode d'exploration interactive d'images en mouvement. Ce dispositif permettra de contrôler le montage de séquences cinématographiques en manipulant un jeu vidéo, offrant ainsi la possibilité d'explorer de manière ludique différentes images cinématographiques par le biais des interactions avec un autre médium de représentation. Cette approche novatrice permettra d'analyser des similitudes visuelles entre ces deux champs, offrant ainsi de nouvelles perspectives de recherche et de développement pour la reconnaissance de l'image.

Mots clés

analyse d'image, intelligence artificielle, cinéma, jeu vidéo, computer vision, visualisation des données, segmentation d'image, jeux de données (*datasets*), image en mouvement, transmédia

1 – Problématique

1.1 – Crises des dispositifs

Situé entre jeu cinéma et jeu vidéo, ce projet de recherche prend comme point de départ théorique les analyses des chercheurs en médiologie, Bolter et Grusin. Dans leur ouvrage consacré au concept de « remédiation¹ » (1999), ils décrivent les diverses stratégies employées par les nouveaux médias pour intégrer et surpasser les codes et contenus des médias « historiques » (précédents). Selon les deux auteurs, il y aurait un double mouvement dans une telle processus de *re*-médiation :

1. Le nouveau médium tente de montrer sa supériorité technique et esthétique pour représenter le monde en intégrant les formes de représentation de ces prédécesseurs.
2. Par ce même besoin d'intégration, on observe l'influence constitutive de l'ancien sur le nouveau. Pour démontrer la qualité de sa représentation, le nouveau média finit par mimer la logique ou le fonctionnement de l'ancien médium qui, par ce même mouvement, fixe la jauge de qualité. L'ancien devient le garant de la nouveauté du nouveau.

Ils expliquent ainsi que : « Les médias visuels numériques peuvent être mieux compris à travers la manière dont ils honorent, rivalisent et révisent la peinture, la photographie, le cinéma, la télévision et l'imprimé en perspective linéaire. [...] Ce qui est nouveau dans les nouveaux médias vient de la manière particulière dont ils remodelent les anciens médias et de la manière dont ils se remodelent eux-mêmes pour répondre aux défis des nouveaux médias². »

Nous proposons d'appliquer cette analyse à la manière dont les jeux vidéo prolongent les médias qui l'ont précédé — en particulier le médium du cinéma narratif hollywoodien classique. À plusieurs égards, le jeu vidéo serait aujourd'hui le médium culturel le plus en pointe. Ne serait-ce que sur le plan économique, l'écosystème du jeu vidéo a dépassé il y a fort longtemps celui du cinéma et de la musique réunis^{3 4}. Comme n'importe quel médium, le jeu vidéo a développé ses propres codes fonctionnels et esthétiques. Grâce à sa capacité à faire évoluer sa narration de façon modulaire et d'intégrer en temps réel des interactions, il tend à se distinguer de la forme linéaire et historiquement située du cinéma.

Pourtant, toute une branche des jeux vidéo — par exemple les jeux à grand budget produits par Sony (*Last Of Us*, *God of War*, *Uncharted*) — continuent de mimer la logique et les thèmes du dispositif cinématographique, jusqu'au point de se couper parfois de sa caractéristique première — l'interactivité. Avec l'arrivée de moteurs de jeux de plus en plus « réalistes » comme Unreal 5

¹ Jay David Bolter, Richard Grusin, *Remediation: Understanding New Media*, Cambridge (MA), The MIT Press, 1999.

² *Ibid.*, p.14-15.

³ 56.5 milliards aux États-Unis seulement pour l'industrie du jeu vidéo, selon le « 2023 Essential Facts About the U.S. Video Game Industry ». Voir : Entertainment Software Association, <https://www.theesa.com/2023-essential-facts>

⁴ « Global Box Office Notched 27% Gain in 2022 to Hit \$26 Billion Total, Research Shows », *Variety*, <https://variety.com/2023/data/news/global-box-office-in-2022-1235480594>

(2022)⁵, il devient toutefois possible de jouer à l'intérieur de séquences parfaitement « linéaires », mais avec la possibilité de réorienter la caméra ou même d'interagir avec des personnages, objets, et décors programmés à l'avance.

1.2 – Crises de la narration

Dans un article au titre provocateur — « *Video games are better without stories* » (2017) — le théoricien des jeux vidéo Ian Bogost a interrogé le besoin des jeux de s'approcher du langage narratif du cinéma : « Le cinéma, la télévision et la littérature racontent tous de meilleures histoires. Alors pourquoi les jeux sont-ils toujours obsédés par la narration ? Une réponse pourrait être l'envie du cinéma. L'industrie du jeu a depuis longtemps rêvé de surpasser Hollywood pour devenir le “médium du 21^e siècle”, un concept désormais si rétrograde qu'il ne pourrait satisfaire qu'un adepte du 20^e siècle⁶. »

Il n'y aurait pas de meilleure illustration de cette envie du cinéma et du jeu vidéo d'épouser leurs codes respectifs que la série *The Last of Us* (HBO, 2023) — une adaptation télévisuelle d'un jeu célèbre (2013–), lui-même inspiré de la série *Planète Terre* (BBC, 2006). Ce jeu, dans sa forme originale, épousait déjà de près les codes, la structure, et l'ambiance d'un film d'action plutôt classique. Le même studio vidéoludique, Naughty Dog, est également à l'origine de la série de jeux *Uncharted* dont l'adaptation cinématographique s'est révélée décevante — probablement parce que ces derniers étaient déjà un pastiche des films d'action *Indiana Jones* de Steven Spielberg.

1.3 – Crises de la représentation

Pour diverses raisons historiques, les ambitions esthétiques du jeu vidéo ont toujours été entravées par la nature de ses configurations technologiques. Dans un article intitulé « Artifactual Playground⁷ », le requérant de ce projet a exploré comment un langage visuel spécifique au jeu vidéo a émergé à partir de la fin des années 1950 en réponse aux contraintes imposées par les limites techniques du médium. Cette situation a généré une ironie fondatrice de l'histoire des jeux vidéo : le désir de simuler des interactions infiniment complexes s'est heurté aux ressources de calcul les plus modestes. Cette contradiction a donné naissance à une période intermédiaire, où l'aspiration au « réalisme » dans les jeux informatiques devait patienter jusqu'à ce que les moyens technologiques soient à la hauteur. Cette anomalie historique a eu un effet « heureux » en favorisant

⁵ Cf. ce tutoriel pour le moteur de jeu Unreal 5 :

<https://dev.epicgames.com/community/learning/tutorials/7Plb/how-quickly-i-create-3d-photorealistic-environments-in-unreal-engine-5>

⁶ Ian Bogost, « Video Games Are Better Without Stories », *The Atlantic*, avril 2017, <https://www.theatlantic.com/technology/archive/2017/04/video-games-stories/524148>

⁷ Douglas Edric Stanley, « On Simulation, Aesthetics, and Play : Artifactual Playground », *Creativeapplications.net*, 2012, <https://www.creativeapplications.net/games/on-simulation-aesthetics-and-play-artifactual-playground>

l'émergence d'une nouvelle esthétique propre au médium, en grande partie influencée par ces limitations.

Aujourd'hui, les moteurs de jeu vidéo sont devenus très puissants dans leur capacité à produire des images ressemblant à celles captées par une caméra de cinéma. On assiste même à de plus en plus de moyens de productions hybrides — par exemple la série *The Mandalorian* (Disney+, 2019) — qui utilisent des moteurs de jeu comme *Unreal 5* pour intégrer directement sur plateau des effets spéciaux. Malgré ces avancées technologiques, le jeu vidéo continue à envier le cinéma et sa façon particulière de capter le réel, y compris lorsqu'il s'agit d'imagerie fantastique. Il va jusqu'à introduire des effets de dispositif spécifiques au médium cinématographiques, comme par exemple le bruit sur la pellicule (*Cuphead*, 2017⁸), le « *motion blur* » (*Mirror's Edge*, 2008 ; *Sonic X*, 2003⁹), ou les vibrations d'une caméra portée à l'épaule (*Gears of War*, 2006 ; *Uncharted*, 2007¹⁰).

1.4 – Problématique

Partant de ces différentes crises et dynamiques de mimétisme entre cinéma et jeu vidéo, ce projet de recherche vise à ouvrir de nouvelles approches dans l'analyse des images en mouvement en approfondissant les liens entre les deux médiums. Dans un premier temps, il mobilise des technologies émergentes de l'intelligence artificielle (IA) pour faciliter la (dé)composition et la catégorisation d'images issues de la création cinématographique et vidéoludique. Dans un deuxième temps, cette analyse sert de base à un travail de recomposition artistique à partir des associations ainsi générées.

Les capacités « génératives¹¹ » des IA ont largement été démontrées ces derniers temps — à travers les *chatbots*, générateurs d'images, et de nouveaux outils qui, peu à peu, commencent à générer des animations (« Frame Interpolation » de Runway ML¹², Zeroscope¹³, EbSynth¹⁴, Kaiber¹⁵), voire même des courtes vidéos plutôt réalistes. Du côté de l'analyse de l'image via les IA, de nombreux outils existent (Open Computer Vision Library¹⁶, Segment Anything Model¹⁷, YOLO¹⁸, Computer Vision Model Zoo¹⁹) ; c'est même leur capacité à analyser et à décomposer divers composants d'une image

⁸ « The 1930's Toons That Inspired The Art of Cuphead », <https://www.ceros.com/inspire/originals/cuphead-game-art>

⁹ « Ten Game With Beautiful Motion Blur », <https://gamingbolt.com/ten-games-with-beautiful-motion-blur/6>

¹⁰ « The different ways screen shake is used in games », <https://www.youtube.com/watch?v=9WBg-KOisUA>

¹¹ Stuart Russel, Peter Norvig, *Artificial Intelligence: A Modern Approach, Global Edition*, Pearson, 2021. Section « 22.7: Unsupervised Learning and Transfer Learning », p. 831.

¹² Frame Interpolation, Runway ML, <https://runwayml.com/ai-magic-tools/frame-interpolation/>

¹³ Zeroscope Text-To-Video, <https://huggingface.co/spaces/fffiloni/zeroscope>

¹⁴ EbSynth, <https://ebsynth.com/>

¹⁵ Kaiber, <https://kaiber.ai/>

¹⁶ OpenCV Foundation, <https://opencv.org/>. Pour un exemple des fonctionnalités d'analyse, voir https://docs.opencv.org/4.8.0/d7/dbd/group_imgproc.html

¹⁷ Segment Anything Model (SAM), Meta, <https://segment-anything.com/>

¹⁸ YOLO: Real-Time Object Detection, <https://pipedie.com/darknet/yolo/>

¹⁹ Nvidia Computer Vision Model Zoo, https://docs.nvidia.com/tao/tao-toolkit/text/model_zoo/overview.html

qui permet aux modèles génératifs de fonctionner. Pourtant, cette analyse reste en grande partie liée à des catégories usuelles du type : « voici un chat, voici une plante, il y a trois piétons dans cette image, voici un chapeau posé sur un pastèque ». Cette approche instrumentale²⁰ a pour revers de limiter l'approche créative car elle reste enfermée dans des catégories déjà établies. En effet, si l'IA se révèle performante pour automatiser des médiums déjà existants (cinéma, jeu vidéo, peinture, littérature, etc.), elle ne dit finalement pas grand-chose des relations profondes qui se tissent à l'intérieur et au-delà de ces champs.

En s'appuyant sur une exploration plus ouverte du découpage conceptuel et thématique d'une image, ce projet de recherche posera les questions suivantes. Comment une démarche artistique mobilisant des technologies d'IA émergentes permet-elle de comprendre ce que le cinéma et le jeu vidéo se doivent et peuvent apprendre l'un de l'autre ? Que peut dire l'IA des catégories habituelles caractérisant les médiums du jeu vidéo et du cinéma ? Comment jouer avec des images en mouvement ?

2 – État de l'art

2.1 – Évolution des formes cinématographiques du collage

Collage Film

Un aspect de ce projet prolonge une pratique cinématographique vieille de presque cent ans d'histoire : la pratique du « *collage film* », appelé aussi selon le contexte « *found footage* », où artistes et cinéastes s'approprient divers extraits de films en tout genre (narrations classiques hollywood, films industriels, *home movies*) et construisent un nouveau film via différentes techniques de montage. Une des premières instances de cette approche est le film *Rose Hobart* (1936)^{21 22}, où l'artiste américain Joseph Cornell remonte le film *East of Borneo* (1931) pour mettre au centre du nouveau film l'actrice Rose Hobart. On trouve un écho du même procédé dans le film du cinéaste autrichien Peter Tscherkassky et son film expérimental *Outer Space* (1991)²³, où Tscherkassky superimpose plusieurs scènes du film *The Entity* de 1980 pour finir avec une sorte de rêverie onirique inquiétante où l'image d'une femme possédée par un démon déborde le corps de l'actrice et commence à hanter le film lui-même. D'autres exemples de *collage film* que nous pourrions citer

²⁰ Anthony Masure, *Design sous artifice : la création au risque du machine learning*, Genève, HEAD – Publishing, coll. « Manifestes », 2023, <https://www.anthonymasure.com/essai-design-sous-artifice>

²¹ Willis Holly, *Rose Hobart*, National Film Registry, 1936, https://www.loc.gov/static/programs/national-film-preservation-board/documents/rose_hobart.pdf

²² Derek Long, *Remixing Rose Hobart*. Dans : *[in]Transition, the Journal of Videographic Film and Moving Image Studies* 5.1, 2018, <https://vimeo.com/225587129>

²³ Tscherkassky, Peter. *Outer Space*. 1991. Lien Mubi : <https://mubi.com/en/us/films/outer-space>

dans cette tradition sont *A Movie* (1958)²⁴ de Bruce Conner, *Our Century* (1982)²⁵ d'Artavazd Peleshian, *Tribulation 99: Alien Anomalies Under America* (1991) de Craig Baldwin²⁶, ou les œuvres de Christian Marclay²⁷.

Avec l'arrivée du Web dans les années 1990, puis les plateformes de partage vidéo comme Bittorrent ou YouTube dans les années 2000, on voit une transformation de ce *collage cinéma* vers dans une nouvelle culture émergente qui prône une logique de « *remix*²⁸ ». Dans ce « *remix culture* », toutes sortes de formes médiatiques peuvent être recombinaées pour faire *œuvre*. Depuis quelques années, le département Cinéma de la HEAD-Genève organise des projets autour d'un concept de *Net Found Footage* et explore la manière dont cette matière protéiforme puisse être recombinaée en de nouvelles formes cinématographiques. En 2021, le département a montré ces travaux au *Festival International de Films de Genève* (GIFF)²⁹.

Cinéma algorithmique

Il existe une autre variante de *remix* cinématographique qui se sert d'algorithmes pour déterminer des re-juxtapositions des séquences d'images. Nous pourrions citer dans cette veine l'installation *Soft Cinema*³⁰ du théoricien Lev Manovich, montrée à l'exposition *Future Cinema* en 2002 au ZKM, Karlsruhe³¹ où des images cinématographiques pouvaient être remontées dans de nouvelles configurations grâce à une base de données qui les re-organisaient dans des associations sémantiques via mots-clés. En 2004, le Centre Pompidou a commandé au requérant de ce projet une installation nommée *The Signal*, où 10 000 séquences de film issues de films industriels américains des années 1950 pouvaient être reconfigurées par les gestes des visiteurs en suivant un modèle sémantique évolutif³². Dans une approche plus proche du *critical media*, le collectif *Disnovation* a réalisé en 2012 l'installation *The Pirate Cinema*³³, qui composait un film infini à partir des séquences d'images partagées en temps-réel via le protocole d'échange de fichiers *Bittorrent*. Enfin, nous pourrions citer l'œuvre *Dérives* d'Émilie Brout et Maxime Marion où l'algorithme

²⁴ Bruce Conner, *A Movie*, 1958, https://en.wikipedia.org/wiki/A_Movie

²⁵ Artavazd Peleshian, *Our Century*, 1982, <https://www.youtube.com/watch?v=BEwdu4azBjA>

²⁶ Craig Baldwin, *Tribulation 99: Alien Anomalies Under America*, 1991.

https://en.wikipedia.org/wiki/Tribulation_99:_Alien_Anomalies_Under_America

²⁷ Christian Marclay, *The Clock*, 2010. [https://en.wikipedia.org/wiki/The_Clock_\(2010_film\)](https://en.wikipedia.org/wiki/The_Clock_(2010_film))

²⁸ Henry Jenkins, Sam Ford, Joshua Green, *Spreadable Media: Creating Value and Meaning in a Networked Culture*, NYU Press, 2011.

²⁹ HEAD – Genève, département Cinéma, *Net Found Footage*, 2022,

<https://www.hesge.ch/head/projet/departement-cinema-au-giff>

³⁰ Lev Manovich, *Soft Cinema*, 2002, <http://www.medienkunstnetz.de/works/soft-cinema>

³¹ Peter Weibel, Jeffrey Shaw, *Future Cinema The Cinematic Imaginary After Film*, MIT Press, 2003,

<https://zkm.de/en/exhibition/2002/11/future-cinema>

³² Douglas Edric Stanley, *The Signal*, Centre George Pompidou, 2004, <https://abstractmachine.net/en/posts/the-signal>

³³ Disnovation, *The Pirate Cinema*, 2012, <https://disnovation.org/piratecinema.php>

construit éternellement un nouveau film à partir de 2 000 séquences de film qui traitent d'une manière ou d'une autre l'eau³⁴.

Jeux cinématographiques

Notre projet cherche à confronter cette pratique du collage — y compris ses variantes algorithmiques plus récentes — avec l'évolution des vingt dernières années du jeu vidéo qui tente également à sa manière de pratiquer l'art du collage cinématographiques, en intégrant les codes visuels, les cadrages, voire même les principes de visionnage du dispositif. Comme mentionné dans notre introduction, il existe de nombreux exemples de jeux qui intègrent divers aspects empruntés du dispositif cinématographique et de son histoire. Des jeux *blockbuster* comme *Lara Croft* (2013), *Uncharted* (2007; 2009; 2011; 2016; 2017), *Red Dead Redemption* (2010; 2018), ou *The Last Of Us* (2013; 2020), empruntent sans ambiguïté aucune les sujets, les plans, et même les *artefacts* du médium cinématographiques. Le boucle semble même se boucler complètement avec la série télévisée *The Last of Us* (2023), basée sur les jeux conçus au départ pour la console de jeu Playstation 4.

Plutôt que de chercher à résoudre ces différentes crises historiques, ce projet propose d'explorer encore plus en profondeur la richesse des interactions possibles entre différents dispositifs de l'image en mouvement, et puis de regarder comment ils sont transformés par les nouveaux outils émergents — en particulier les intelligences artificielles.

2.2 – *Inside Inside* : une installation artistique à la base de ce projet

Ce projet de recherche prend comme point de départ un dispositif très particulier — l'installation *Inside Inside*³⁵ — qui tente de confronter tous ces enjeux que nous venons d'évoquer. Cette installation, développée par le requérant du présent projet, et a été exposée entre autres au GIFF en 2021³⁶. *Inside Inside* (images en annexe) est une installation interactive qui mêle jeux vidéo et cinéma grâce à un réseau de neurones. Dans ce dispositif, les joueuses contrôlent un personnage central — un petit garçon — qui tente de survivre dans la dystopie du jeu vidéo populaire *Inside*. En parallèle, un réseau de neurones analyse en temps réel la séance du jeu Playstation et tente de reconnaître « à l'intérieur » d'autres images à partir d'une liste d'images issues de dystopies cinématographiques.

Au fur et à mesure que la joueuse évolue d'un niveau à l'autre, le système analyse l'environnement du jeu, ainsi que d'autres facteurs, et reconnaît certaines fonctionnalités. Il associe ensuite ces images à des images sur le deuxième écran, extraites d'une base de données de séquences organisées de l'histoire du cinéma. En conséquence, les joueur·euses peuvent essentiellement

³⁴ Émilie Brout, Maxime Marion, *Dérives*, 2011-2014. <https://www.eb-mm.net/en/projects/derives>

³⁵ *Inside Inside*, infos du projet : <https://abstractmachine.net/en/posts/inside-inside> & <https://vimeo.com/589844238>

³⁶ *Inside Inside*, Genève, GIFF, 2021, <https://www.giff.ch/archives/2021/programme/virtualterritories/index.html>

re-séquencer ces images en s'attardant, avançant, reculant ou sautant tout au long du jeu via la sélection intégrée de « chapitres » du jeu : le *timing*, l'ordre et la durée dépendent entièrement de la nature du *gameplay* de la personne qui joue. Alors que les joueur·euses évoluent dans l'espace du *gameplay*, ils·elles montent simultanément un film.

En s'inspirant de ce système développé pour le dispositif *Inside Inside*, nous proposons de porter plus loin plusieurs pistes très prometteuses qui ont été découvertes dans ce travail. Notamment :

- Le développement plus en profondeur d'une plateforme interactive où l'on peut naviguer à l'intérieur des médias historiques par le biais d'une interface ludique.
- Le perfectionnement d'un système qui permet de cataloguer divers objets, artefacts visuelles, et éléments formels dans une image en mouvement par le biais des intelligences artificielles.
- La création de nouveaux formats transmédia, par exemple des documentaires interactifs où l'on peut jouer avec une image en temps réel, tout en déclenchant des informations historiques ou critiques de cette même image.

2.3 – État de l'art technique

Ce projet de recherche s'appuie fortement sur les diverses techniques d'analyse de l'image qui évoluent actuellement à toute vitesse grâce aux IA, mais se base également sur des techniques plus anciennes, issues principalement de la bibliothèque de développement OpenCV³⁷. Le projet usera donc de techniques à la fois émergentes mais aussi anciennes dans son modèle d'analyse de l'image. Au lieu de s'appuyer sur des plateformes de génération grand public comme OpenAI, Runway, ou Midjourney, ce projet propose de revenir aux sous-couches plus fondamentales en s'appuyant sur l'histoire riche de la bibliothèque OpenCV. Cette bibliothèque peut non seulement importer les modèles les plus actuels, mais contient également l'étendue de toutes les techniques historiques de la « vision machinique » les plus importantes, datant parfois, pour certains algorithmes, des années 1960. Par exemple, pour détecter le mouvement à l'intérieur d'une image — c'est-à-dire de l'image dans son ensemble et reconnaître s'il s'agit d'un travelling latéral ou en avant — l'utilisation de la fonction « *optical flow* » d'OpenCV serait adaptée. Enfin, un autre avantage d'OpenCV est de pouvoir fonctionner non seulement sur des machines performantes avec des cartes graphiques Nvidia, mais également sur des plateformes plus légères comme les Raspberry PI ou les Jetson Nano.

Le système technique derrière l'œuvre *Inside Inside* s'appuie sur un type de réseau de neurones appelé « *Convolutional Neural Network* » (CNN), qui a « appris » à la machine à reconnaître les contours des formes dans le jeu et dans les séquences de films. Depuis, d'autres techniques ont été développées concernant l'identification des images dans leur ensemble (le film, le jeu, la séquence)

³⁷ Open Computer Vision Library, <https://opencv.org>

mais aussi les éléments à l'intérieur de l'image (les objets, les personnages, les lieux, etc.). Dans le cadre de ce projet de recherche, une des pistes à explorer serait de développer de nouvelles formes d'analyse de l'image à partir des techniques de segmentation en plein développement grâce au nouveau « *Segment Anything Model*³⁸ » (SAM) publié par Meta Research³⁹ (avril 2023), et déjà intégré dans des plateformes de labellisation comme Roboflow⁴⁰. Une telle avancée⁴¹, non seulement dans la reconnaissance de l'image mais aussi dans les outils de labellisation de celles-ci, rendrait beaucoup plus réaliste le travail titanesque d'apprentissage des images et sous-images dans les jeux et les films que ce projet devra réaliser.

3 – Objectifs

Les techniques d'IA proposées ici pour l'analyse de l'image ont comme origine des usages industriels liés au commerce et au profilage d'individus ou d'objets – soit donc un cadre conceptuel « extractiviste⁴² » à l'intersection de la police et du capitalisme, et qui renforce les pouvoirs et discriminations déjà en place⁴³. Ce projet propose, au contraire, de mobiliser ces technologies dans un contexte artistique afin de contribuer à une réflexion esthétique et pratique croisant cinéma et jeu vidéo.

Objectif général

– Développer un modèle d'entraînement et d'inférence d'images issu des histoires du cinéma cinématographique et du jeu vidéo. Ce modèle inclura des techniques d'analyse de l'image comme la segmentation sémantique (objets, personnages, décors), ainsi que des notions structurales (ex. : cadrage, éclairage, mouvement, etc.).

Objectifs analytiques

– Réfléchir aux spécificités et frontières des médiums du jeu vidéo et du cinéma.
– Superposer en temps réel des formes de commentaires audio ou d'analyse graphique sur les films et les jeux. Il s'agirait ici de concevoir une sorte de piste supplémentaire de lecture des médias (comme les commentaires des DVD) qui puisse s'adapter à l'image en fonction des interactions.

Objectifs créatifs

³⁸ <https://segment-anything.com>

³⁹ Alexander Kirillov *et al.*, « Segment Anything », 2023, <https://arxiv.org/abs/2304.02643>

⁴⁰ « Roboflow Annotate », <https://blog.roboflow.com/how-to-use-segment-anything-model-sam>

⁴¹ Pour une présentation de l'état d'art de ces techniques, https://www.youtube.com/watch?v=X4laF0_oOdc

⁴² Crawford, Kate. *Contre-atlas de l'intelligence artificielle. Les coûts politiques, sociaux et environnementaux de l'IA* [2021], trad. de l'anglais (Australie) par Laurent Bury, Paris, Zulma, 2022.

⁴³ Pasquinelli, Matteo. *The Eye of the Master: A Social History of Artificial Intelligence*, Londres, Verso Books, 2023.

- Reconnaître des sous-éléments dans des images pour ouvrir de nouvelles possibilités d'interaction, qu'il s'agisse d'images conçues pour être interactives (jeu vidéo) ou non (cinéma).
- Inventer de nouvelles formes de média hybrides, où l'exploration à l'intérieur d'un jeu peut déclencher des séquences d'autres médias et ainsi approfondir le récit.
- Explorer des approches plus expérimentales de l'image, comme par exemple utiliser les mouvements de la caméra ou des personnages d'un film, manipuler des déplacements à l'intérieur d'un jeu, ou reconstruire un lieu de tournage.

4 – Méthodologie

Partant de l'expérience acquise par le requérant sur son œuvre *Inside Inside* (2021), et plus généralement sur sa connaissance historique et technique des champs de l'intelligence artificielle et du jeu vidéo (articles, cours, projets artistiques, etc.), ce projet mobilise une démarche de recherche-crédation hybridant esthétique, programmation et installation. Plus précisément, conçu dans une approche exploratoire pour être mené dans une durée de 12 mois, il se focalisera sur la production d'un modèle d'analyse d'image, à savoir un modèle d'entraînement permettant de traiter en temps réel des séquences de cinéma et de jeu vidéo.

Les « modèles » de *machine learning* comportent deux phases : « l'entraînement » et « l'inférence ». Ces deux phases peuvent être résumées comme « l'avant » et « l'après » du modèle. La phase d'entraînement concerne la phase d'apprentissage des séquences de film et de jeux, et l'identification à l'intérieur de celles-ci des objets, personnages, décors, et autres éléments liés à sa composition. L'inférence est la phase interactive, où le modèle est capable de regarder un jeu ou un film et d'identifier en temps réel les divers aspects étiquetés dans la phase entraînement.

Afin de développer et d'éprouver un modèle d'entraînement et d'inférence, nous produirons :

- 1 – Un *dataset* d'entraînement d'images de films issus de l'histoire du cinéma (westerns), couplé avec un autre *dataset* d'entraînement d'images issus d'un jeu vidéo (*Red Dead Redemption*, Rockstar Games, 2010-2018). Le choix de cette thématique en amont du démarrage du projet permettra de se focaliser sur les nombreux défis techniques.
- 2 – Une installation et/ou site Web permettant d'explorer et de mettre en scène les deux types de *datasets*.
- 3 – Un nouveau format d'audioguide interactif pour les images en mouvement, offrant la capacité d'adapter en temps réel un commentaire audio au contenu visuel affiché à l'écran.

4 – Planification et résultats

Durée : 12 mois (janvier à décembre 2024).

Certaines phases sont menées en parallèle afin de respecter le délai envisagé.

Workpackage #1 – Curation (2 mois : janvier à mars 2024)

Description – Recherche de séquences cinématographiques à étiqueter, principalement issues de films de western produits entre 1930 et 1990. Création d’une base de données de séquences (références TMDb⁴⁴, *timecodes* avec durée, premières indications de formes à identifier dans les images).

Livrable – Base de données des 100 films sélectionnées, avec *timecodes* des séquences.

Workpackage #2 – Étiquetage (6 mois : février à juillet 2024)

Description – Importation des séquences cataloguées pendant la phase de curation dans l’outil d’étiquetage. Étiquetage des séquences avec les mots clés identifiés pendant la phase curation.

Livrable – *Dataset* d’étiquetage sur une plateforme en ligne.

Workpackage #3 – Développement (6 mois : mai à septembre 2024)

Description – Développement du modèle d’entraînement et d’inférence des images. Affinage du modèle en fonction des résultats d’entraînements.

Livrables – Un modèle d’entraînement ; une séquence (cahier Jupyter) d’entraînement du modèle ; une séquence (cahier Jupyter) d’inférence du modèle ; une API d’inférence du modèle.

Workpackage #4 – Intégration (2 mois : octobre à novembre 2024)

Description – Création d’une installation interactive à partir du travail fourni dans les phases des *workpackages* #2 et #3.

Livrables – Dispositif interactif ; publication des codes sources sur la plateforme GitHub.

Workpackage #5 – Documentation (2 mois : novembre à décembre 2024)

Description – Documentation du processus, de la phase de la curation jusqu’à l’intégration. Documentation de l’API. Documentation du modèle d’entraînement de l’inférence.

Livrable – Documentation placée sur la plateforme GitHub.

⁴⁴ The Movie Database, <https://www.themoviedb.org>

5 – Valorisation

5.1 – Contribution à des plateformes *open source*

Ce projet s'appuie sur (et contribue à) la communauté *open source* de recherche et de partage des « modèles » et « *datasets* » en *machine learning*. Les deux plateformes les plus importantes de cette communauté sont Huggingface⁴⁵ et Kaggle⁴⁶. Il existe également des plateformes importantes de partage d'étiquetage, comme CVAT⁴⁷ et Roboflow⁴⁸. Ce projet publiera trois types de données – annotations, modèles et *datasets* – sur le site de Huggingface. L'étiquetage sera conçu directement sur le site de Roboflow afin de partager la méthode et les résultats au plus grand nombre. La création de *datasets* à partir de séquences de films et de jeux sera la phase la plus chronophage du projet et représentera donc une contribution importante pour la communauté. Une copie des données créées sera placée dans une plateforme *open data* « FAIR » comme Zenodo ou DaSCH.

5.2 – Article de recherche

Un article synthétique du projet sera publié dans la revue *Issue* (HEAD – Genève)⁴⁹, avec des liens d'accès vers les outils de documentation développés.

5.3 – Communication externe

Après la clôture administrative du projet (et sur le budget de la HEAD), un communiqué de presse sera rédigé par un·e journaliste. Une vidéo de valorisation sera réalisée en interne et mise en ligne sur le YouTube HEAD afin de démontrer le fonctionnement du dispositif⁵⁰. Ces deux items seront adressés à la communauté scientifique via des newsletters spécialisées, ainsi qu'à la communauté professionnelle via la newsletter de la HEAD et des posts sur les médias sociaux, notamment LinkedIn.

5.4 – Présentation du projet

L'installation et/ou le site Web produit dans le cadre du projet pourront être présentés dans des événements et festivals tels que le festival Kikk, le Geneva International Film Festival, le Game Developers Conference, ou dans des revues spécialisées telles que *Back Office*, *Creative Applications*, ou *Neural*.

⁴⁵ Huggingface, <https://huggingface.com>

⁴⁶ Kaggle, <https://www.kaggle.com>

⁴⁷ Computer Vision Annotation Tool (CVAT), <https://www.cvat.ai>

⁴⁸ Roboflow, <https://roboflow.com>

⁴⁹ *Issue Journal*, <https://issue-journal.ch>

⁵⁰ Exemple de vidéo de valorisation : « TheKniTGeekResearch », 2021, <https://youtu.be/Y3A6AlunfkC>

6 – Équipe et compétences

6.1 – Requéran

Bio – Douglas Edric Stanley est maître d'enseignement au Master Media Design de la HEAD – Genève (HES-SO), où il enseigne le design algorithmique, les médias ludiques et prend en charge l'innovation pédagogique. Entre 1998 et 2020, il était professeur d'arts numériques à l'École supérieure d'art d'Aix en Provence, où il a fondé L'atelier hypermédia, un atelier qui traite l'algorithme et le code en tant que matières plastiques avec un accent particulier sur le rôle du jeu dans l'exploration des formes émergentes. Il a donné de multiples workshops sur la programmation artistique pour diverses associations, musées, universités et écoles d'art. Il a participé en tant que curateur et artiste à de nombreuses expositions liées à l'art informatique. Au sein de la HEAD – Genève, il participe ou a participé aux projets de recherche « Design et *machine learning* » (dir. Anthony Masure, 2021-2023), « Fashion Narratives » (dir. Alexia Mathieu, 2022-2023).

Site Web : <http://abstractmachine.net>

Rôle – Douglas Edric Stanley, responsable du projet, participera au développement du modèle, de son intégration, et de la direction artistique du projet dans son ensemble. Chargé de coordonner l'équipe du projet, il rendra compte des heures de travail effectuées, organisera des points de suivi réguliers avec le responsable de recherche de la HEAD – Genève, et respectera les délais prévus.

6.2 – Équipe

Deux autres rôles seront attribuées à des collaborateur·trices choisies via des fiches de postes :

1 – Une personne responsable du développement du modèle. Cette personne travaillera à la fois sur les phases d'entraînement (l'apprentissage) du modèle et d'inférence (l'interaction) avec le modèle. Elle travaillera étroitement avec le requérant du projet, développera avec lui l'architecture du modèle, son interface d'entraînement, et aidera à la mise en place de la chaîne de travail sur des serveurs dédiés, principalement dans des environnement virtuels et *datacenters* en ligne.

2 – Une personne responsable de l'étiquetage des images de cinéma et de jeux vidéo. Cette personne aura des compétences dans le domaine de l'analyse des images cinématographiques et du jeu vidéo. Elle organisera avec le requérant principal du projet une base de données d'images de films et de jeux vidéo, et procédera au travail d'identification des composants principaux des images.

7 – Ressources

- Bellour, Raymond. *L'analyse du film*. Albatros, 1980.
- Bellour, Raymond. *L'Entre-Images: Photo, Cinéma, Vidéo*. La Différence, 1990.
- Blanchet, Alexis. « Cinéma et jeux vidéo : trente ans de liaisons ». *MédiaMorphoses*, n° 22, 2008, p. 33-38.
- Blanchet, Alexis. *Les jeux vidéo au cinéma*. Armand Collin. 2012.
- Blanchet, Alexis. *Des Pixels à Hollywood*. Éditions Pix'n'Love. 2010.
- Boidy, Maxime. *Les études visuelles*. Presses universitaires de Vincennes, 2017.
- Boyer, Elsa. *Le conflit des perceptions*. MF, 2015.
- Bolter, Jay David; Grusin, Richard. *Remediation: Understanding New Media*. The MIT Press, 2000.
- Bordwell, David ; Thompson, Kristin. *Film Art: An Introduction*. McGraw-Hill, 2009.
- Burdick, Anne; Drucker, Johanna; Lunenfeld, Peter; Presner, Todd; Schnapp, Jeffrey. *Digital Humanities*. The MIT Press, 2016.
- Conner, Bruce. *A Movie*. Film. 1958. Lien *Archive.org* : <https://archive.org/details/a-movie-dir-bruce-conner-1958>.
- Cook, David A. *A History of Narrative Film*. 3rd ed., illustrated. W.W. Norton, 1996.
- Crawford, Kate. *Contre-atlas de l'intelligence artificielle. Les coûts politiques, sociaux et environnementaux de l'IA* [2021], trad. de l'anglais (Australie) par Laurent Bury. Zulma, 2022.
- Deleuze, Gilles. *L'Image-mouvement : Cinéma 1*. Minuit, 1983.
- Deleuze, Gilles. *L'Image-temps : Cinéma 2*. Minuit, 1985.
- Drucker, Johanna. *Graphesis, Visual Forms of Knowledge Production*. Harvard Press, 2014.
- Ehmann, Antje ; Eshun, Kodwo (eds.). *Harun Farocki Against What? Against Whom?* Walther König, 2010.
- Farocki, Harun; Ujică, Andrei. *Videograms of a Revolution*, 1992. Film documentaire. <https://www.imdb.com/title/tt0108489/>
- Fourmentraux, Jean-Paul. *Des images performatives et opératoires : Autour de Jean-Luc Moulène*. Études photographiques. 2017.
- Guylas, Aaron. *The Paranormal and the Paranoid: Conspiratorial Science Fiction Television*. Rowman & Littlefield, 2015.
- Heller-Nicolas, Alexandra. *Found footage horror films : fear and the appearance of reality*. McFarland and Co., 2014.
- Klevan, Andrew; Clayton, Alex; eds. *The Language and Style of Film Criticism*. Routledge, 2011.
- Kirillov, Alexander, et al. *Segment Anything*. 2023. <https://arxiv.org/abs/2304.02643>
- Jenkins, Henry; Ford, Sam; Green, Joshua. *Spreadable Media: Creating Value and Meaning in a Networked Culture*. NYU Press. 2011.
- Manovich, Lev. *Soft Cinema*, 2002, <http://www.medienkunstnetz.de/works/soft-cinema>
- Manovich, Lev. *The Language of New Media*. The MIT Press, 2002.

- Manovich, Lev. *Software Takes Command*. Bloomsbury Academic, 2013.
- Manovich, Lev. *Cultural Analytics*. The MIT Press, 2020.
- Masure, Anthony. *Design sous artifice : la création au risque du machine learning*. HEAD – Publishing, 2023.
- Metz, Christian. *Essais sur la signification au cinéma*. Paris, Klincksieck, 1968.
- Pasquinelli Matteo. *The Eye of the Master: A Social History of Artificial Intelligence*. Verso Books, 2023.
- Perret, Catherine. *Politique de l'archive et rhétorique des images*. Critique, 2010.
- Rose, Gillian. *Visual Methodologies*. SAGE. 2001.
- Russel, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. Random House. 2020.
- Russell, Stuart; Norvig, Peter. *Artificial Intelligence: A Modern Approach*, Global Edition, 4th ed. Pearson. 2021.
- Saint-Jevin, Alexandre. « Essai pour une méthode d'analyse plastique du vidéoludique, Vers une approche esthétique-psychanalytique des jeux vidéo ? ». *Open Edition*, 2018.
- Szeliski, Richard. *Computer Vision: Algorithms and Applications*, 2nd edition. The University of Washington, 2020.

Annexe – Images du dispositif *Inside Inside* (2021)

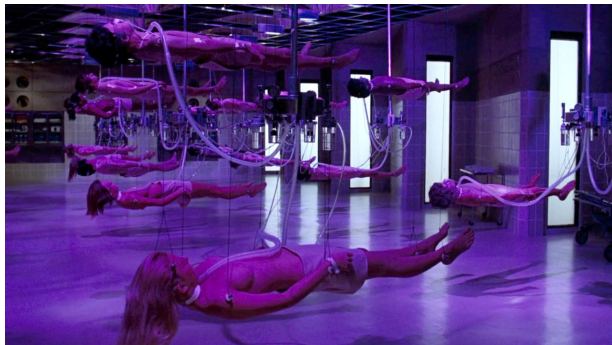




E.T. The Extra-Terrestrial, Stephen Spielberg, 1982 & *Inside*, Playdead, 2016



Artificial Intelligence, Steven Spielberg, 2001 & *Inside*, Playdead, 2016



Coma, Michael Crichton, 1978 & *Inside*, Playdead, 2016



Brainstorm, Douglas Trumbull, 1983 & *Inside*, Playdead, 2016



Deux images du procédé d'étiquetage du dispositif *Inside Inside*, avant sa phase d'entraînement. Ce sont ces composants qui sont recherchés par le programme pendant la phase d'inférence en temps réel avec le jeu.