

SUJET DE PROJET TECHNIQUE DE FIN DE E3

2^e SEMESTRE 2022/2023

NOMS DES ÉLÈVES (4 minimum obligatoire) :

1. 2. 3. 4.

TITRE DU PROJET : à trouver

MOTS-CLÉS : Application, vidéo, IA, Traitement d'images.

DESCRIPTION DU PROJET :

Pour un photographe, effacer en un clic un élément inesthétique de sa photo (comme une personne, une poubelle, une voiture...) était impossible il y a encore quelques années, mais cela est désormais possible avec l'IA. Cela produit un rendu parfait / indécélable (même avec des arrière plans assez complexes). Exemple de logiciels pour cela : SnapEdit, Photoshop...



En revanche, effacer un objet sur une **vidéo** est incommensurablement plus compliqué ! D'une part, car une vidéo est composée d'environ 24 photos par secondes, et d'autre part, car sur chaque image, chaque partie effacée/recomplétée par l'IA, doit être parfaitement cohérente avec les images précédentes et suivantes au cours du temps, en fonction des mouvements de la caméra. Sans quoi l'on verrait des « artefacts » visuels survenir. C'est d'une complexité mathématique folle pour pouvoir être réalisé de manière satisfaisante/indécélable.

Si bien qu'actuellement, même les meilleurs logiciels de retouche vidéo produisent un résultat imparfait / décelable à l'œil nu. Cela crée une légère zone de trouble, de pixelisation, de légers clignotements (*flickering*) autour de l'objet en mouvement (en mouvement relatif par rapport à la caméra, que ce soit lui et/ou la caméra qui bouge).

Or, des chercheurs américains ont réussi récemment à mettre au point un algorithme d'effacement d'objets dans une vidéo, basé sur l'IA, qui surpasse tous les meilleurs logiciels actuels. Cet outil porte le nom barbare de « *End-to-End Framework for Flow-Guided Video Inpainting* », en abrégé « E2FGVI » (« Cadre de bout en bout pour la repeinture vidéo guidée par le débit ») et produit un résultat indécélable ! Le résultat est proprement époustouflant. Aperçu :

<https://www.youtube.com/watch?v=N--qC3T2wc4>

Fait incroyable : cet algorithme a été laissé en open source par ses créateurs !

Les divers tutoriels d'installation et ses lignes de codes sont disponibles au lien suivant :

<https://github.com/MCG-NKU/E2FGVI>

Il y a un lien « Quickstart » qui semble permettre de lancer l'algorithme :

<https://colab.research.google.com/drive/12rwY2gtG8jVWINx9pjmmM8uGmh5ue18G>

L'objectif est évidemment d'exécuter correctement cet algorithme, mais aussi de créer une interface qui soit ergonomique et utilisable par un profane de l'informatique.

TRAVAIL À RÉALISER :

1) **OUVERTURE DE LA VIDÉO** : ouvrir un fichier vidéo (d'à peu près n'importe quel format, mais principalement : mp4, mkv, mov, wmv, avi, avchd, webm), et prenant en charge les principaux codecs (h264, h265, apple prores...) et de n'importe quelle résolution (jusqu'à des images 4K en 3840x2160 pixels, voire pourquoi pas 6K...)

2) **DÉLIMITATION DE L'OBJET À EFFACER** : à l'aide d'un outil de pinceau, l'utilisateur doit pouvoir délimiter à main levée, assez grossièrement, le contour de la zone à effacer de la vidéo (le plus souvent une personne ou un objet). Cette étape produit une zone sélectionnée que l'on appelle un "masque".

3) **TRACKING VIDÉO + EFFACEMENT + INPAINTING** : par la suite, l'utilisateur pourra cliquer sur un bouton, qui devra déclencher la procédure suivante (qui peut être assez longue) :

- un suivi constant (dit "tracking" vidéo) de la zone délimitée par le masque, au cours du temps. C'est un processus qui va, image par image, faire évoluer la forme de ce masque pour qu'il s'adapte à chaque instant aux contours de l'objet, qui varieront plus ou moins au cours du temps, selon ses mouvements dans le cadre. Après que l'utilisateur ait délimité l'objet à retirer sur une frame de référence de son choix, il pourra par la suite déclencher le tracking automatique de l'objet via un bouton adapté. Ce tracking pourra s'effectuer soit en avant, soit en arrière à partir de cette image, au choix. Une fois que le logiciel aura dessiné le masque sur chaque frame, il devra pouvoir être laissée la possibilité à l'utilisateur, en cas d'erreur de tracking, de corriger le masque sur une certaine plage de temps. Pour ce faire, il suffira à l'utilisateur d'utiliser l'outil pinceau pour ajuster la forme du masque (ajouter ou retirer des zones), puis de relancer le processus de tracking (en avant ou en arrière).

- un effacement de tous les pixels recouverts par le masque, sur chaque image.

- un remplissage, avec de nouveaux pixels générés par l'IA, de la zone effacée (étape qu'on appelle "inpainting" vidéo). Bien évidemment, cette image de remplissage doit être très cohérente avec le reste de l'image, et cohérente également avec toutes les autres images avoisinantes temporellement, pour éviter les artefacts vidéo (pixellisations, flous, clignotements, lignes anormales, distorsions de type "chewing gum"...).

4) [s'il reste du temps] **AIDE DE L'IA PAR L'UTILISATEUR** : il se peut parfois que le programme (mais cela est censé être vraiment rare) peine à produire un résultat très satisfaisant. Cela peut arriver lorsque la zone effacée est assez étendue et/ou lorsqu'elle recouvre des objets très complexes, que l'IA n'est pas en mesure de reconstituer, comme le tableau de la Joconde par exemple (si ce tableau n'est pas apparu suffisamment jusque là dans la vidéo). Si c'est le cas, on peut envisager une quatrième étape facultative qui permettrait à l'utilisateur de "guider" l'IA en lui soumettant l'image à laquelle devrait globalement ressembler la zone cachée dans la première frame effacée. On appelle usuellement une telle image, le plus souvent fabriquée de toutes pièces par un artiste graphique sur Photoshop, une "image de référence". C'est une fonctionnalité utilisée chez beaucoup de logiciels professionnels, mais elle sera peut-être complexe à mettre en oeuvre.

5) **EXPORTATION** : le programme devra enfin permettre d'exporter le projet en un nouveau fichier vidéo, sans avoir dégradé la qualité du fichier original (que ce soit en terme de résolution, de netteté, de couleurs, de compression...), et idéalement en pouvant choisir un format et un codec de sortie.

Il est demandé de porter une attention particulière à la **durée** de l'étape 3 ! Pour traiter par exemple 1 minute de vidéo, combien de temps cela prendrait-il environ avec un PC spécial montage vidéo ? Et avec un Mac Book Pro ? (avec la puce M1). Ce serait un problème si cela devait nécessiter des jours et des jours de calcul...

OUTILS MATÉRIELS / LOGICIELS SUPPORT : Idéalement, il faudrait que le programme tourne sur plusieurs plateformes, principalement Mac et PC, Windows (et éventuellement Linux). A priori pas sur Android ni iOS.

**URL DU PROJET LE PLUS PROCHE AUQUEL CETTE PROPOSITION DE PROJET PEUT ÊTRE
COMPARÉE :**

<https://www.youtube.com/watch?v=N--qC3T2wc4>

=====
Accord du responsable de projet de fin d'année du département :

Le 20 / 03 / 2023

Note Projets E3 2022-20233