

# SYNTHESE ET SEGMENTATION CARDIAQUES CONTROLEES PAR LE TEXTE : DE LA GENERATION D'IRM 2D A LA MODELISATION VOLUMETRIQUE 3D

Halima FOUADI<sup>A,B</sup>, Mohamed KAS<sup>A</sup>, Yassine RUICHEK<sup>A</sup>, Youssef EL MERABET<sup>B</sup>

<sup>A</sup> Université de Technologie Belfort Montbéliard, UTBM, CIAD UR 7533, F-90000 Belfort, France

<sup>B</sup> Université Ibn Tofail faculté des sciences, UIT-FSK, SETIME, Kenitra, Maroc

L'acquisition de bases de données IRM cardiaques finement annotées reste un défi majeur pour l'apprentissage profond en raison de contraintes éthiques et logistiques. Pour pallier ce manque, nous proposons un cadre génératif unifié permettant la synthèse contrôlée de données cardiaques anatomiquement cohérentes à partir de descriptions textuelles. Cette présentation détaille une approche 2D pour la génération conjointe d'images et de labels, et son extension 3D pour la modélisation de maillages.

Dans un premier temps, le pipeline 2D s'appuie sur *Low-Rank Adaptation* (LoRA) pour générer des masques pathologiques et sur *ControlNet* pour synthétiser les coupes IRM correspondantes. Pour cette étape, nous avons généré uniquement les coupes médianes dans le plan petit axe (*short-axis*). Ces paires synthétiques enrichissent ensuite l'entraînement d'un modèle de segmentation *SegFormer*. Dans un second temps, l'approche s'étend à la génération 3D des maillages avec l'architecture de diffusion texte-vidéo *CogVideoX*. En assimilant les volumes à des séquences temporelles (pseudo-vidéos) et guidé par le *Rectified Flow Matching*, le modèle traduit des indicateurs cliniques structurés en représentations géométriques volumétriques plausibles.

L'évaluation 2D démontre un réalisme visuel avec un *Fréchet Inception Distance* (FID) de 1,0640. L'ajout de ces données à l'entraînement *SegFormer* améliore la segmentation (Dice moyen de 93,00 % sur ACDC) et prouve une excellente robustesse au changement de domaine (vers la cohorte M&Ms). L'extension 3D maintient cette haute fidélité : les maillages générés respectent les variations morphologiques des pathologies avec un score de plausibilité anatomique de 0,9626 et un 3D-FID de 21,96.

En conclusion, la transition d'une synthèse 2D vers une modélisation 3D contrôlée par le texte établit une passerelle novatrice entre langage clinique et anatomie, ouvrant de vastes perspectives pour l'augmentation de données en imagerie cardiovasculaire