



TÉLÉCOM PARIS

PROJET IMA 206

Variational Auto-Encoders For Cardiac Shape Modeling

Yanis Aït El Cadi
Zakaria Chahboune
Clément Leroy
Tanguy Schmidtlin

Encadré par
Loïc Le Folgoc

24 juin 2024

Résumé

Ce projet explore les applications des modèles statistiques de forme en imagerie médicale, notamment dans la description de la variabilité anatomique et l'amélioration de la segmentation. Il vise à établir une correspondance entre un espace latent de forme cardiaque et des masques de segmentation 2D, en utilisant des auto-encodeurs variationnels. L'objectif est de comprendre comment cet espace latent encode l'anatomie cardiaque, en examinant quelles variables latentes capturent des caractéristiques telles que la taille du cœur ou l'épaisseur du myocarde. Des analyses sont menées pour déterminer si les sous-groupes pathologiques sont distincts dans cet espace.

Table des matières

1	Description du modèle	1
1.1	Données disponibles	1
1.2	Pré-traitement du dataset	1
1.3	Architecture du modèle	2
1.4	Fonction de coût	2
2	Evaluation du modèle	4
2.1	Evaluation des performances pour la reconstruction des images	4
2.1.1	Indice de Sørensen-Dice	4
2.1.2	Résultats	4
2.2	Evaluation des performances pour la génération de nouvelles images	5
2.2.1	Fréchet Inception Distance (FID)	5
2.2.2	Résultats	6
2.3	Recherche des hyperparamètres optimaux	6
3	Analyse de l'espace latent	10
3.1	Etude des informations encodées le long d'une dimension de l'espace latent	10
3.2	Visualisation de l'espace latent à l'aide d'UMAPs	14
3.2.1	Vérification de l'overfit	14
3.2.2	Identification des outliers	14
3.2.3	Position dans l'espace latent des points associés à des hauteurs différentes	15
3.2.4	Position dans l'espace latent des points associés à des pathologies différentes	16
4	Conclusion	18

Chapitre 1

Description du modèle

Cette partie décrit la base de donnée utilisée et l’architecture du modèle d’auto-encodeur variationnel implémenté.

1.1 Données disponibles

Les données utilisées dans de ce projet sont un ensemble d’images du coeur de 150 patients [1]. Les images correspondent à des coupes horizontales du torax. Un ensemble de coupes forme une image 3D du coeur à un instant donné. Plusieurs images 3D capturant des instants différents permettent de rendre compte de l’action du coeur. Chaque image est accompagnée d’une segmentation qui distingue les quatre zones suivantes : le ventricule droit, le ventricule gauche, le myocarde, et le fond qui correspond à la zone extérieur au coeur.

Les patients sont divisés en classes dans lesquels sont regroupés les patients présentant la même pathologie cardiaque. Le diagnostic est effectué par des medecins compétents. Les images sont accompagnées de métadonnées concernant les patients étudiées, elles comprennent la taille du patient, son poids et son groupe (sa pathologie).

Les groupes sont les suivants :

- DCM (dilated cardiomyopathy)
- HCM (hypertrophic cardiomyopathy)
- MINF (previous myocardial infarction)
- NOR (normal subjects)
- RV (abnormal right ventricle)

1.2 Pré-traitement du dataset

Les images cardiaques des patientes ont été extraites. Seules les images 3D correspondant au maximum de dillatation et au maximum de contraction ont été conservées. Les métadonnées des images ont également été conservées. Chaque image 3D est composée d’une dizaine de coupes, chaque coupe correspondant à une hauteur différente du coeur.

Les coupes ont été recadrées afin d’éliminer une partie du fond puis ont été redimensionnées en images de 128 par 128 pixels afin de travailler avec des images de tailles semblables. Le redimensionnement a été calculé en utilisant la méthode du plus proche voisin afin de conserver des valeurs cohérentes pour les segmentations. De plus, les coupes d’une même image 3D sont recadrées selon la même bounding box dont la taille est choisie comme étant la plus petite qui peut contenir toutes les zones segmentées du coeur sur toutes les

tranches. Ainsi, les coupes aux extrémités du coeur contiennent des zones de coeur plus étroites que les coupes du milieu du coeur.

Initialement, chaque image contient un canal de couleur pouvant prendre l’une des quatre valeurs suivantes : 0 pour le fond, 1 pour le ventricule droit, 2 pour le myocarde et 3 pour le ventricule gauche. Nous convertissons ce canal en quatre canaux distincts. Cela permettra au modèle de facilement produire en sortie les probabilités d’appartenance aux quatre zones considérées pour chaque pixel.

Finalement, le jeu de données est passé dans un dataloader afin de pouvoir utiliser des lots de données (batches) pendant l’entraînement.”

1.3 Architecture du modèle

Nous avons implémenté un β Variational Auto-Encoder (β -VAE) semblable à ceux déjà expérimentés dans la littérature [2] [3]. Nous décrivons brièvement l’architecture retenue.

Encodeur : L’encodeur est une série de couches convolutives suivies de couches entièrement connectées. Il commence par plusieurs blocs convolutifs qui réduisent progressivement la taille spatiale de l’image d’entrée tout en augmentant le nombre de canaux. Chaque bloc convolutif comprend deux couches de convolution avec des fonctions d’activation ELU. Les tailles des blocs convolutifs sont successivement : (4, 48), (48, 96), (96, 192), puis (192, 384). À la fin, les données sont aplaties et passent par deux couches linéaires pour produire les paramètres de la distribution latente : les moyennes μ et les log-variances \logvar .

Décodeur : Le décodeur commence par une couche entièrement connectée qui prend le vecteur latent comme entrée et produit une carte de caractéristiques de taille $384 \times 8 \times 8$. Cette carte de caractéristiques passe ensuite par plusieurs blocs de convolution transposée pour augmenter progressivement la taille spatiale et réduire le nombre de canaux. Chaque bloc de déconvolution comprend une couche de convolution transposée suivie d’une convolution avec des fonctions d’activation ELU. Les tailles des blocs convolutifs sont successivement : (384, 192), (192, 96), (96, 48) puis (48, 4).

1.4 Fonction de coût

La fonction de coût utilisée classiquement lors de l’entraînement du β -VAE combine l’entropie croisée à la divergence de Kullback-Leibler pour assurer un équilibre entre la reconstruction et la génération. Voici l’expression de la loss comme exprimée dans [2] :

$$\mathcal{L} = \mathbb{E}_{q_{\theta_e}(\vec{z}|\mathbf{x})}[-\log q_{\theta_d}(\mathbf{x}|\vec{z})] + \text{KL}(p_{\theta_e}(\vec{z}|\mathbf{x}) \parallel p(\vec{z})), \quad (1.1)$$

où $\vec{z} \in \mathbb{R}^N$ est un vecteur latent, θ_e sont les paramètres de l’encodeur, et θ_d sont les paramètres du décodeur. $p_{\theta_e}(\vec{z}|\mathbf{x})$ est une distribution Gaussienne et $q_{\theta_d}(\mathbf{x}|\vec{z})$ est la distribution de \mathbf{x} conditionnée par \vec{z} .

Nous utilisons une fonction de coût qui dépend d’un paramètre β et qui permet de modifier l’équilibre entre les compétences attendues en reconstruction d’image et en génération de nouvelles images. La fonction de coût devient donc :

$$\mathcal{L} = \mathbb{E}_{q_{\theta_e}(\vec{z}|\mathbf{x})}[-\log q_{\theta_d}(\mathbf{x}|\vec{z})] + \beta \text{KL}(p_{\theta_e}(\vec{z}|\mathbf{x}) \parallel p(\vec{z})). \quad (1.2)$$

La Figure 1.1 présente les performances du VAE dont l’architecture a été décrite ci-dessus. Pour cette simulation, les hyperparamètres ont été fixés comme énoncé dans la Table 1.1. Sur les représentations des segmentations, les zones noires correspondent au fond, les zones en gris foncé correspondent au ventricule droit, les zones en gris clair correspondent au myocarde et les zones blanches correspondent au ventricule gauche.

TABLE 1.1 – Paramètres d'entraînement

Dimension de l'espace latent	8
nombre d'epochs	100
taille du batch	8
learning rate	6×10^{-5}
β	100

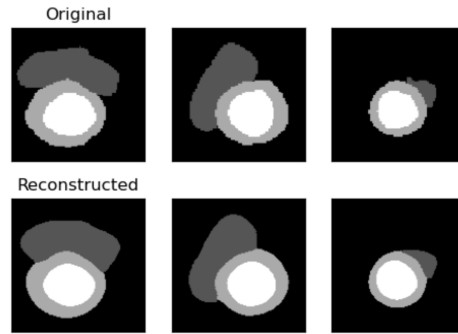


FIGURE 1.1 – Images originales et images reconstruites par le VAE

Chapitre 2

Evaluation du modèle

Le modèle d’auto-encodeur variationnel choisi pour étudier les modèles de coeur dépend de plusieurs hyper-paramètres qui conduisent à des résultats après entraînement différents. Nous explicitons dans cette partie les métriques d’évaluation du modèle qui nous permettent d’estimer les hyperparamètres optimaux.

2.1 Evaluation des performances pour la reconstruction des images

2.1.1 Indice de Sørensen-Dice

Afin de mesurer la performance de notre modèle sur la reconstruction des images, nous utilisons l’indice de Sørensen-Dice qui permet de quantifier la ressemblance entre l’image de segmentation originale et l’image de segmentation reconstruite par le modèle. Plus l’indice est élevé, plus la reconstruction est fidèle.

Les segmentations sont constituées de trois régions d’intérêt : le myocarde (R_1) et les deux ventricules (R_2 - R_3) (le fond est omis ici). Etant donnée une image originale I_{ori} et une image reconstruite I_{rec} , on définit l’indice de Sørensen-Dice associé à la région $R_i \in \{R_1, R_2, R_3\}$ par :

$$SD_{R_i} = \frac{2|I_{ori}(R_i) \cap I_{rec}(R_i)|}{|I_{ori}(R_i)| + |I_{rec}(R_i)|}, \quad (2.1)$$

où $I_{ori}(R_i)$ (resp. $I_{rec}(R_i)$) est l’ensemble des pixels de l’image I_{ori} (resp. I_{rec}) appartenant à la région R_i , est où $|\cdot|$ désigne le cardinal de l’ensemble.

Finalement, la métrique est la moyenne des indices de Sørensen-Dice de chaque région d’intérêt.

2.1.2 Résultats

La figure 2.1 présente trois exemples de reconstructions d’images par un modèle entraîné avec les paramètres décrits dans la table 1.1.

On observe que les images d’origine (données en entrée de l’encodeur) sont pixelisées. Cela est dû à la méthode d’interpolation du plus proche voisin qui introduit un crénelage dans les images lorsque la nouvelle dimension (ici 128×128) est supérieur à la dimension d’origine. On observe que les structures sur les images reconstruites (reçues en sortie du décodeur) ont tendance à être plus arrondies que sur les images d’origine, ce qui fait perdre de l’information. Nous verrons dans la partie 3.1 quelles coordonnées de l’espace latent jouent sur cette particularité.

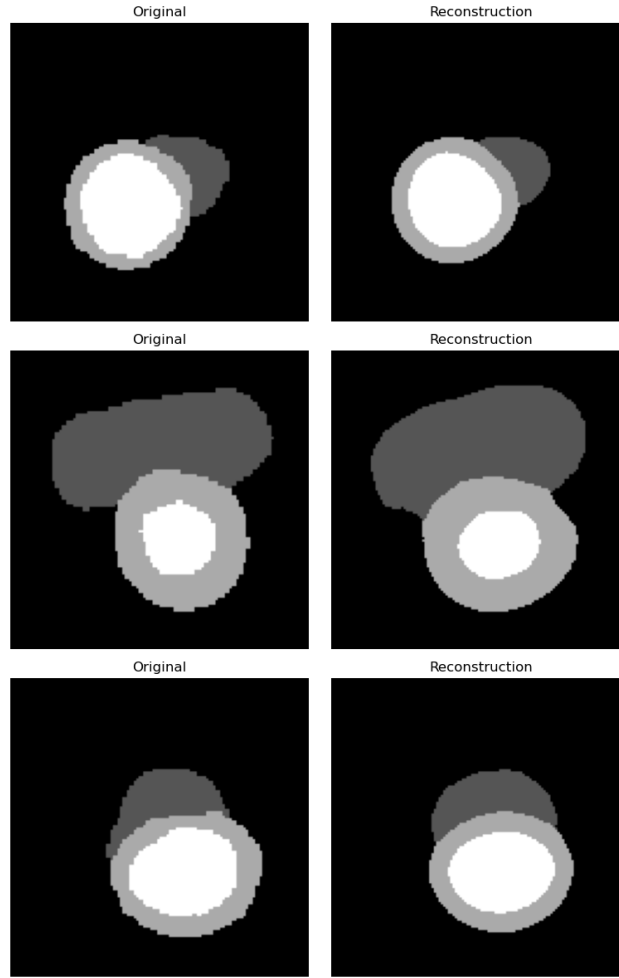


FIGURE 2.1 – Exemples de couples d’images originales (à gauche) et reconstruites (à droite)

2.2 Evaluation des performances pour la génération de nouvelles images

2.2.1 Fréchet Inception Distance (FID)

Pour analyser la qualité de la génération d’images de notre modèle, nous utilisons la distance de Fréchet. La Fréchet Inception Distance (FID) est une métrique couramment utilisée pour évaluer la qualité des images générées par des modèles génératifs, tels que les Generative Adversarial Networks (GANs) ou les VAE comme dans notre cas. Plus la FID est faible, plus les images générées sont considérées comme crédibles, puisqu’elles présentent alors des caractéristiques similaires aux images réelles.

On souhaite calculer la distance de Fréchet entre les distributions des caractéristiques des images réelles et des images générées. Pour obtenir ces distributions, on utilise un réseau de neurones pré-entraîné (Inception v3, ici) pour extraire des caractéristiques des images. Plus précisément, on passe les images à travers ce réseau et on collecte les activations de la dernière couche avant la classification finale, ce qui produit des vecteurs de caractéristiques. Les caractéristiques extraites des images réelles et générées sont supposées suivre une distri-

bution normale multidimensionnelle. Pour chaque ensemble de caractéristiques, on calcule la moyenne (μ) et la matrice de covariance (Σ).

La distance de Fréchet entre deux distributions gaussiennes multivariées μ_1, Σ_1 et (μ_2, Σ_2) est donnée par :

$$\text{FID} = \|\mu_1 - \mu_2\|^2 + \text{Tr}(\Sigma_1 + \Sigma_2 - 2(\Sigma_1 \Sigma_2)^{1/2}) \quad (2.2)$$

où Tr désigne la trace d'une matrice (la somme de ses éléments diagonaux), et $\|\mu_1 - \mu_2\|^2$ est la distance euclidienne au carré entre les vecteurs de moyenne.

Un FID plus bas indique que la distribution des caractéristiques des images générées est plus proche de celle des images réelles, ce qui signifie que les images générées sont de meilleure qualité et plus similaires aux images réelles.

2.2.2 Résultats

Sur la Figure 2.2, nous pouvons voir trois images générées par le modèle en suivant une loi uniforme centrée réduite. Sur les trois images, les structures ont des emplacements et des tailles relatives cohérentes. Sur des échantillons plus large, il est parfois possible d'observer des anomalies, comme un ventricule droit trop grand comparément au reste ou un myocarde trop épais. Il n'y a pas de problème par rapport aux positions relatives du myocarde et du ventricule gauche, le ventricule étant toujours bien à l'intérieur du myocarde. Sur de rares images (environ 0.5% des cas), le ventricule droit est séparé en deux parties, avec l'une des deux partie étant plus petite que l'autre.

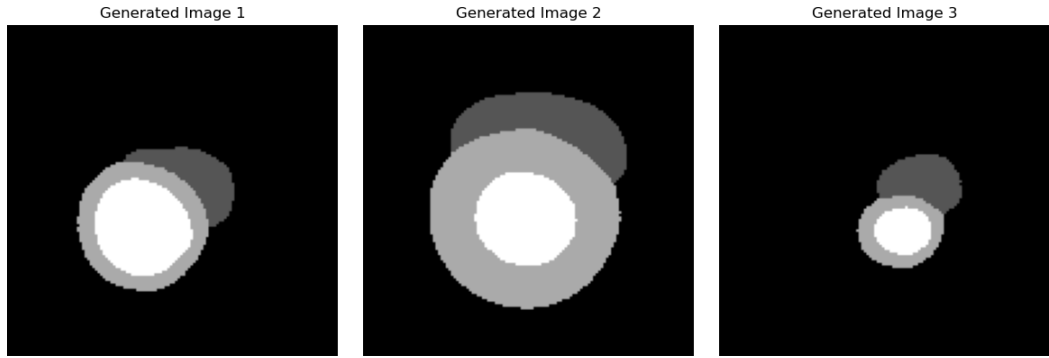


FIGURE 2.2 – Exemple d'images générées par le modèle

2.3 Recherche des hyperparamètres optimaux

Grâce à la FID et à l'indice de Sørensen-Dice, nous pouvons quantifier la qualité d'un modèle entraîné et donc rechercher le meilleur modèle qui maximise chacun de ces deux indicateurs. Pour ce faire nous avons entraîné des modèles avec des paramètres différents et pour chacun de ceux-ci, nous avons calculé la FID et l'indice de Sørensen-Dice associé. Les paramètres testés ainsi que leurs valeurs sont les suivants :

- dimension latente : 4, 8, 12, 16, 32, 64,
- taille du batch : 8, 16, 32,
- valeur de beta : 1, 10, 100, 500, 1000.

Toutes les figures de cette partie sont tracées en utilisant 100 epochs par entraînement.

Les graphes présentés en figure 2.3 montrent les évolutions de nos métriques quand β varie. On observe un fort impact de β sur la FID, variant d'environ 600 quand β est faible à moins de 200 quand β est proche de 1000. Ce résultat est cohérent avec nos attentes

puisque l'objectif d'introduire le paramètre β avec $\beta > 1$ était justement d'inciter le modèle à produire des résultats suivant la distribution $p(\vec{z})$ supposée normale centrée pour l'espace latent. Donc lorsque l'on prend un vecteur aléatoirement selon la loi $\mathcal{N}(0^{\mathbf{K}}, 1^{\mathbf{K}})$ (avec \mathbf{K} la dimension de l'espace latent), le passage par le décodeur permet d'obtenir une image proche de la distribution du dataset. Pour ce qui est du Dice, la valeur de β ne semble pas avoir d'impact significatif.

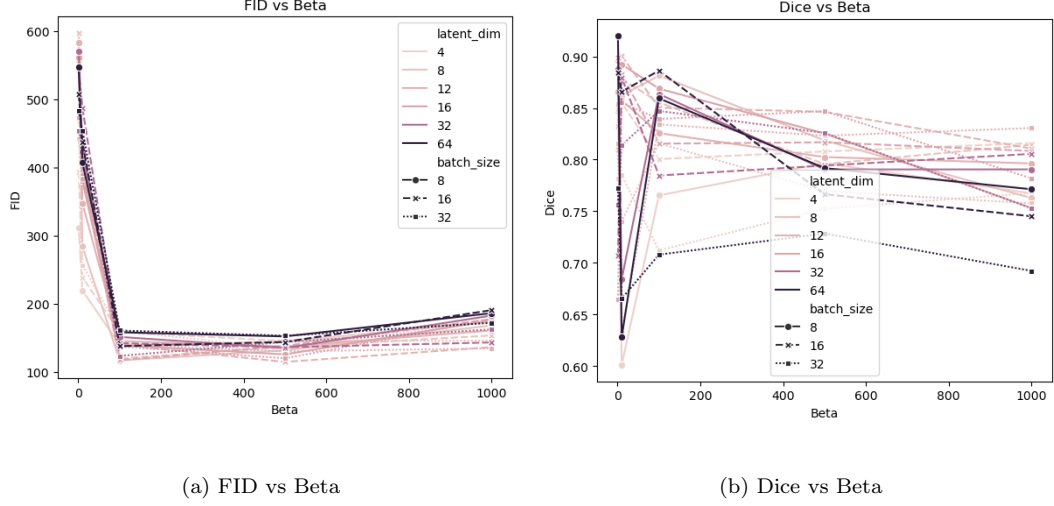
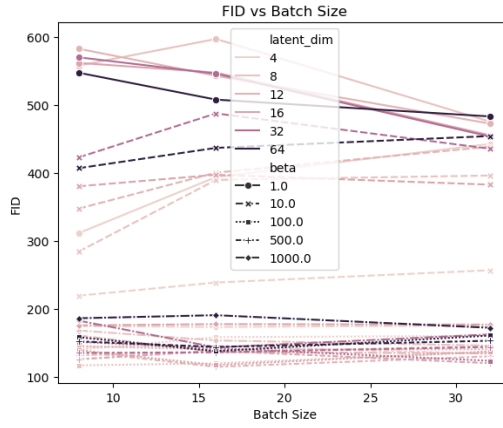


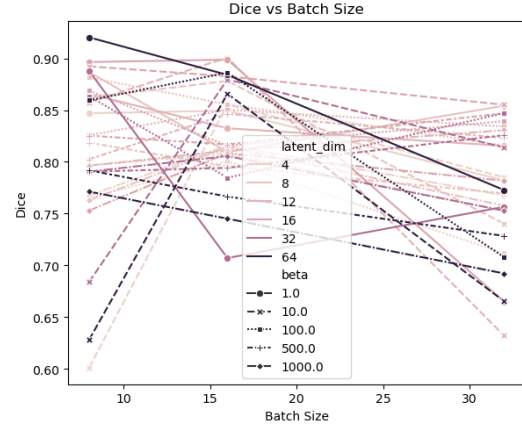
FIGURE 2.3 – Impact de β sur la FID et l'indice de Sørensen-Dice.

La figure 2.4 présente la FID et l'indice de Dice comme des fonctions de la taille du batch. La taille du batch ne semble pas avoir d'impact ni sur le FID ni sur le Dice.

Il est possible que lorsque le modèle converge de manière stable, la modification de la taille du batch n'ait pas d'impact significatif sur nos métriques considérées. En effet, la variabilité des gradients introduite par les différentes taille de batch peut être moyennée au cours de l'entraînement, conduisant ainsi à des résultats semblables. Il est également possible qu'il y ait des plages de valeurs de taille de batch dans lesquelles le modèle fonctionne de manière similaire. Par exemple, peut-être y a-t-il une valeur seuil au-dessus de laquelle la taille de batch n'offre plus d'avantage supplémentaire.



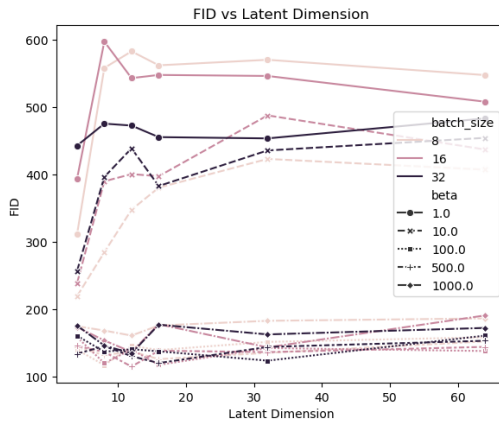
(a) FID vs Batch Size



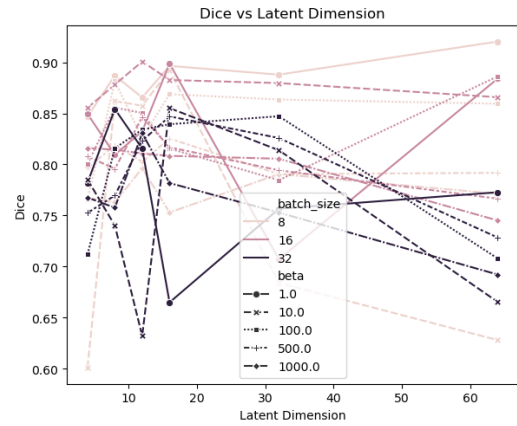
(b) Dice vs Batch Size

FIGURE 2.4 – Impact de la taille du batch sur la FID et l'indice de Sørensen-Dice.

La relation entre la dimension latente et nos métrique est montrée en figure 2.5. La dimension latente ne semble pas avoir d'impact sur la performance de notre modèle (peu d'impact sur la FID et pas d'impact significatif sur l'indice de Dice). Il est possible que le modèle soit capable d'encoder l'ensemble des informations contenues dans les images dans un espace latent de dimension faible. Ainsi, augmenter la taille de l'espace latent ne permettrait pas de mieux encoder l'information et cela expliquerait les résultats observés. Nous pourrions vérifier cette hypothèse en un second temps en étudiant les points dans l'espace latent. Si les points étaient proches de tous appartenir à un sous espace vectoriel de dimension strictement inférieur à celle de l'espace latent, cela prouverait que les dimensions de l'espace latent ne sont pas toutes utilisées pour coder l'information.



(a) FID vs Latent Dimension



(b) Dice vs Latent Dimension

FIGURE 2.5 – Impact de la dimension latente sur la FID et l'indice de Sørensen-Dice.

Les FID et indices de Sørensen-Dice de l'ensemble des modèles sont présentés en figure 2.6. Elle indique que pour la plupart des combinaisons, le modèle performe bien et similairement. On y voit encore le lien entre la valeur de β et le FID ainsi que la disparité dans la valeur que prends le Dice. La disparité des couleurs à travers le graphique tend à montrer que les performances du modèles sont indépendantes de variations de la dimension latente.

Un arbitrage entre performance de reconstruction et de génération doit être fait pour déterminer quels sont les hyperparamètres considérés comme optimaux. Pour la suite, nous considérons les paramètres qui donnent le points le plus en haut à gauche dans la figure 2.6. Ce choix est arbitraire mais traduit grossièrement la maximisation des deux métriques de performances : minimization de la FID (performance de génération) et maximisation de l'indice de Dice (performance de reconstruction).

Les paramètres sont donc les suivants : une dimension latente égale à 8, une taille de batch égale à 8, et un β égal à 100.

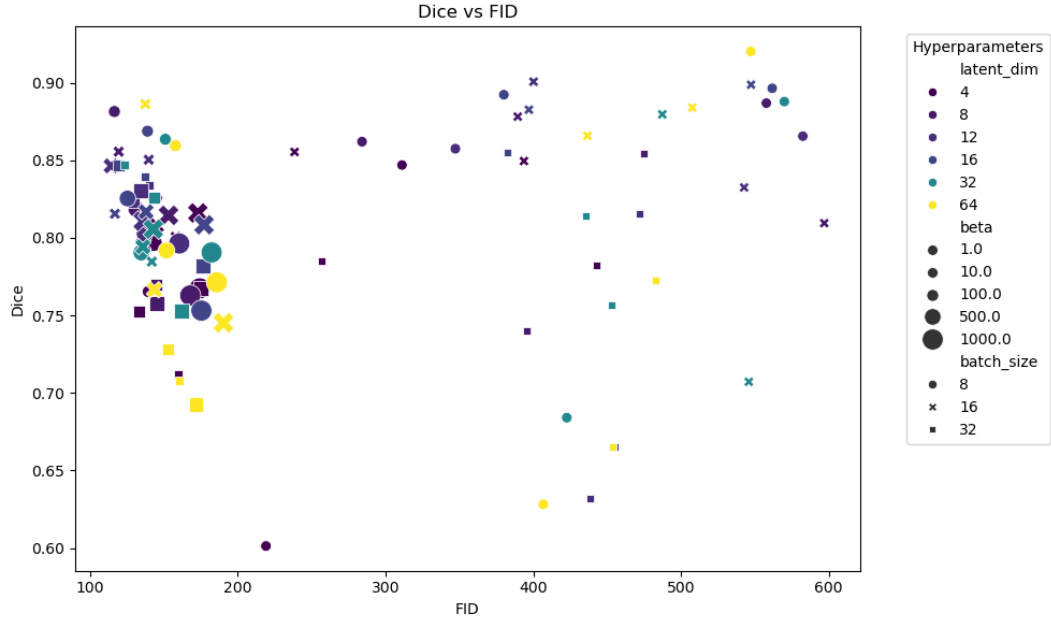


FIGURE 2.6 – FID et indice de Sørensen-Dice pour des modèles avec différents paramètres de taille de batch, de dimension latente et de β .

Chapitre 3

Analyse de l'espace latent

Nous avons défini l'architecture du modèle ainsi que ses hyperparamètres obtenus en maximisant les performances de reconstruction des images et de génération cohérente de nouvelles images. Dans cette partie, nous étudions la répartition des caractéristiques des images dans l'espace latent du VAE.

3.1 Etude des informations encodées le long d'une dimension de l'espace latent

Le VAE encode les différentes caractéristiques du dataset dans l'espace latent. On s'intéresse maintenant au lien entre les caractéristiques des images et la position du point encodé associé dans l'espace latent. L'espace latent étant de dimension 8, il est possible d'analyser chaque dimension indépendamment. Pour ce faire, nous avons encodé l'ensemble des images du dataset puis nous avons observé les intervalles dans lesquels les coordonnées des points obtenus varient. Nous avons échantillonné ces intervalles puis avons placé les points obtenus à l'entrée du décodeur pour obtenir de nouvelles images.

Les Figures 3.1 et 3.2 montrent deux exemples de variations du vecteur latent, chaque rangée correspond à la variation d'une coordonnée du vecteur latent, de la valeur minimale observée dans l'ensemble des données réelles à la valeur maximale observée dans l'ensemble des données réelles.

On observe un impact de la première coordonnée sur la taille et l'orientation de l'ensemble myocarde-ventricule gauche (MVG) - il passe d'allongé verticalement à allongé horizontalement - ainsi que sur la position du ventricule droit qui remonte le long du MVG. La seconde coordonnée joue sur la taille du ventricule droit qui grandit à mesure que la valeur de la coordonnée latente augmente. La troisième coordonnée ne semble pas avoir d'impact visible sur les structures. La quatrième coordonnée semble avoir un rôle similaire à la première car il semblerait que seul le ventricule droit soit affecté en remontant le long du MVG. La cinquième coordonnée impacte la taille du ventricule droit. La sixième coordonnée joue un rôle sur la largeur du ventricule gauche et sur les épaisseurs du myocarde et du ventricule droit. La coordonnée la plus intéressante est la septième, elle encode directement la taille des structures, plus la valeur est faible et plus les structures sont petites, plus la valeur est grande plus les structures sont grandes. On peut faire le parallèle entre la taille des structures et la hauteur de la slice dans l'IRM, cette coordonnée encode donc la hauteur de la slice. La dernière coordonnée est très similaire à la cinquième de part les changements qu'elle montre.

De manière générale, il semble qu'il y ait une certaine redondance entre certaines coordonnées de l'espace latent. Cela appuie l'hypothèse formulée dans la partie 2.3 concernant la relation entre la dimension de l'espace latent et les performances du modèle. Nous avons en effet vu que choisir une dimension latente de 8 plutôt que de 4 ne faisait augmenter le Dice et baisser le FID que légèrement. Il serait donc possible d'effectuer une analyse similaire sur un modèle avec une dimension latente de 4 pour potentiellement obtenir des distinctions plus prononcées entre les différentes coordonnées.

Variations for latent_101

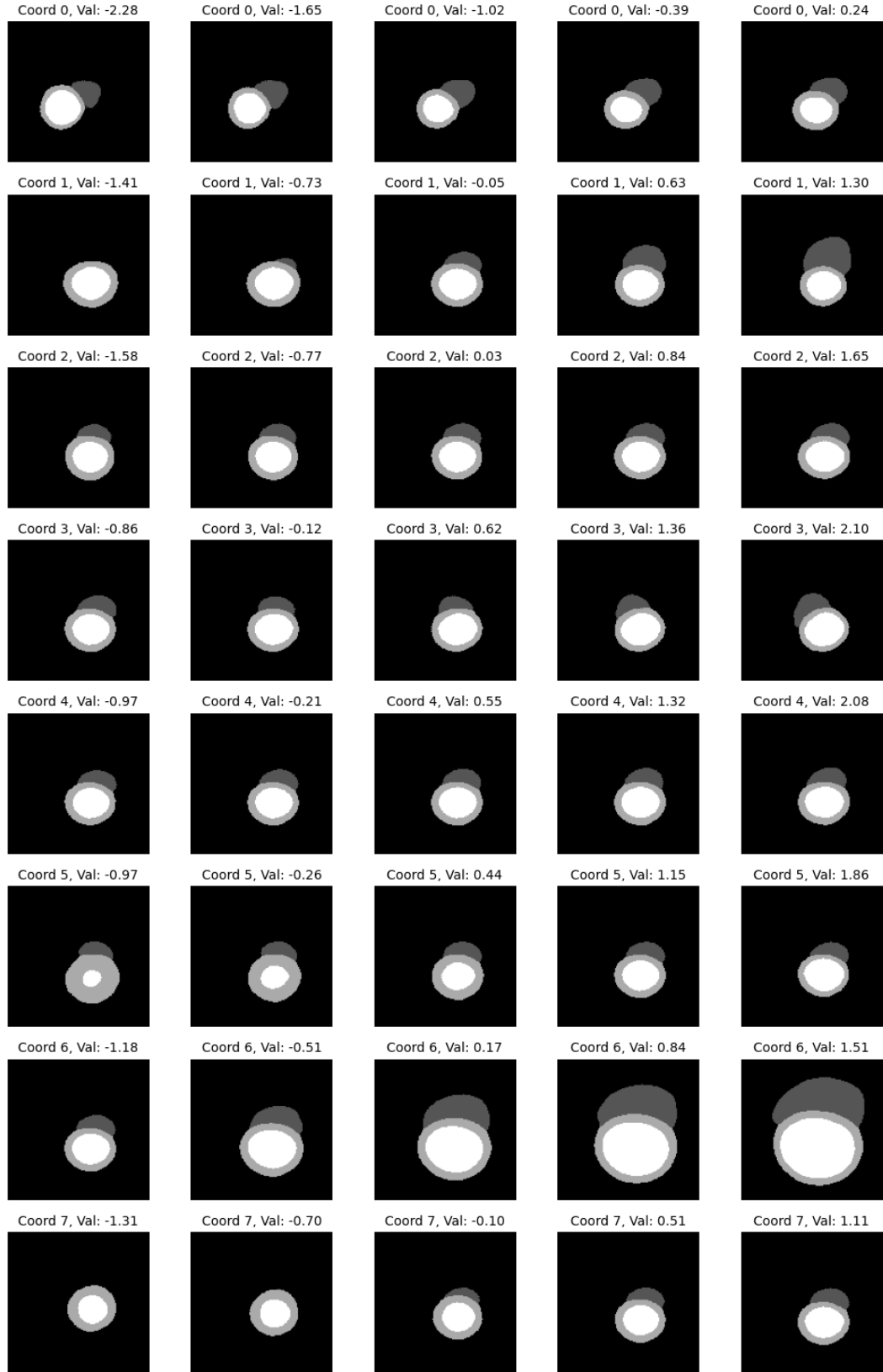


FIGURE 3.1 – Patient 18 : impact des variations du code latent

Variations for latent_5

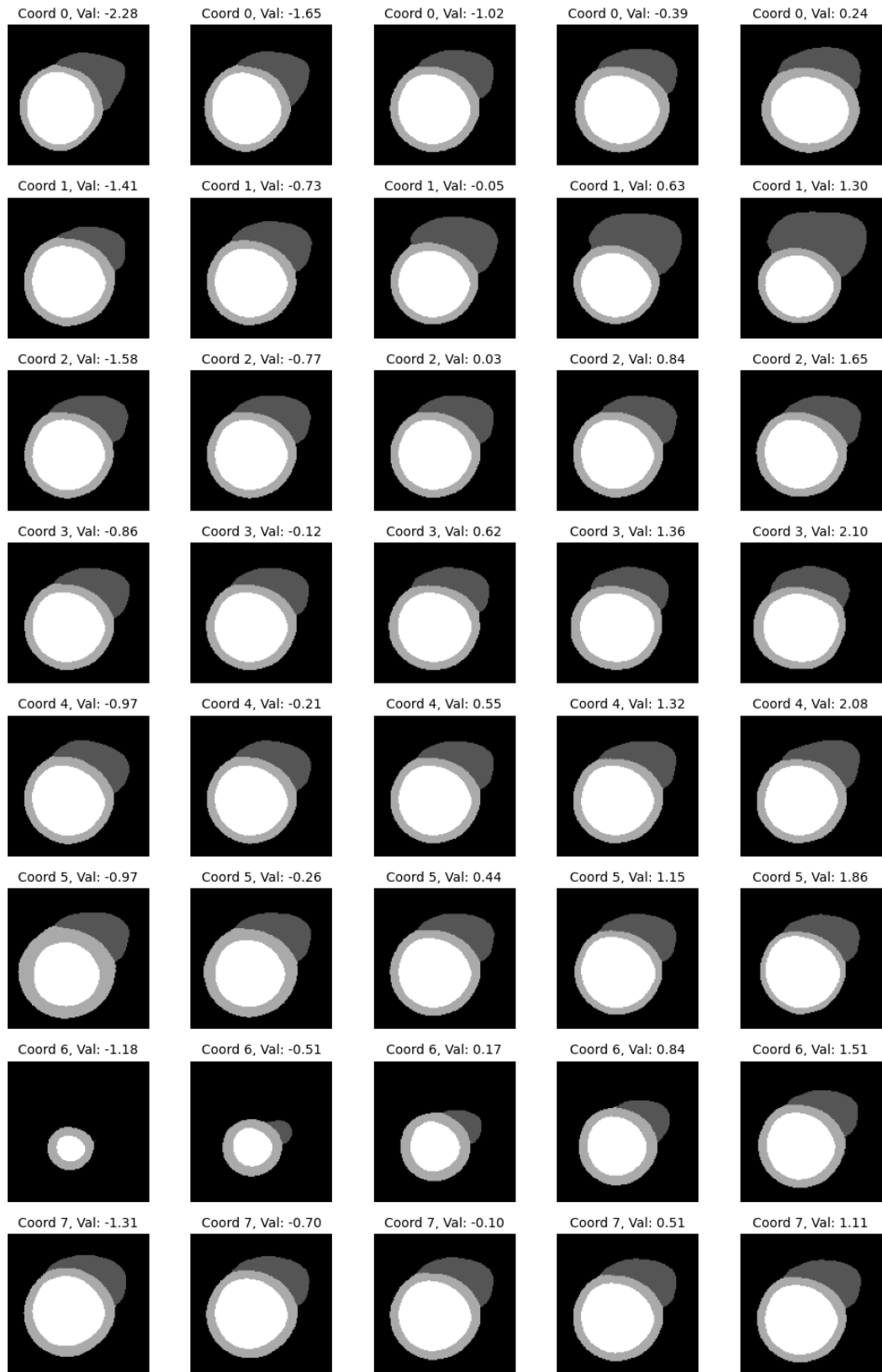


FIGURE 3.2 – Patient 54 : impact des variations du code latent

3.2 Visualisation de l'espace latent à l'aide d'UMAPs

L'espace latent a été projeté sur un espace en dimension 2 qui permet une visualisation simplifiée des points de l'espace. Les points associés à des caractéristiques différentes sont affichés avec des couleurs distinctes. Nous espérons ainsi que les projections permettront de comprendre comment l'espace latent encode l'anatomie cardiaque.

3.2.1 Vérification de l'overfit

Premièrement, nous avons souhaité vérifier l'absence d'overfitting. Pour cela, les données encodées d'entraînement et de test ont été projetées ensemble, comme présenté en Figure 3.3. Visuellement, les données de test sont réparties dans l'espace de la même manière que les données d'entraînement, ce qui indique que le modèle généralise correctement.

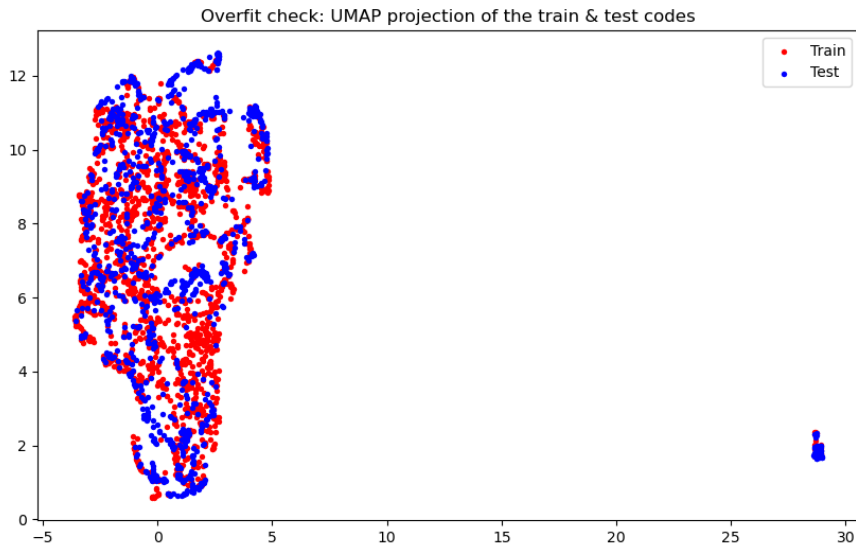


FIGURE 3.3 – UMAP des points associés aux trains et aux tests

3.2.2 Identification des outliers

Nous remarquons aussi la présence d'un petit cluster d'outliers, complètement à l'écart de la masse centrale, que nous avons souhaité identifier. Pour cela, nous avons annoté la hauteur de chaque coupe par rapport au volume 3D du coeur associé, avec 0 pour l'apex bas et 1 pour le haut. De plus, pour certaines images 3D, il y a des coupes au dessus et en dessous du coeur, qui deviennent donc des images vides en 2D, et que nous avons annotées comme tel.

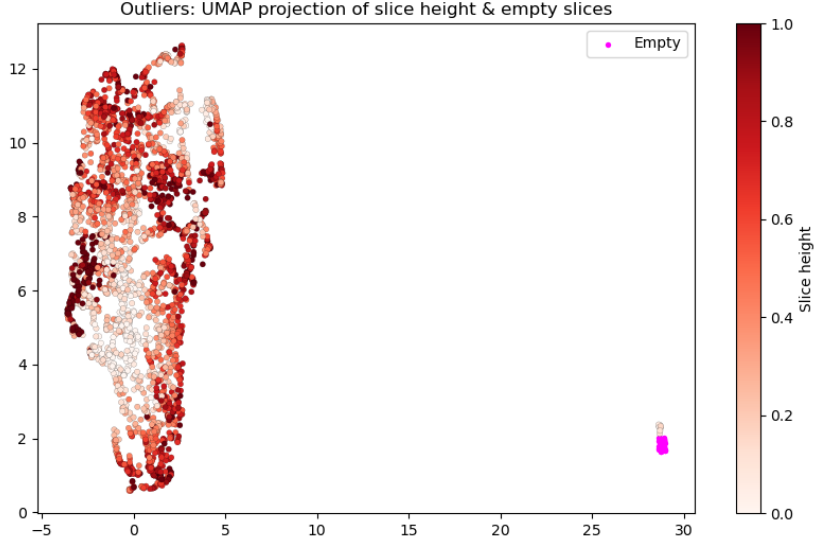


FIGURE 3.4 – Projection des coupes vides et de la hauteur des coupes non-vides

La figure 3.4 indique que ce cluster d’outliers est en réalité l’ensemble de toutes les coupes vides, ainsi que quelques coupes non-vides proches de l’apex bas du coeur. Ces dernières sont présentées sur la figure 3.5, et sont en réalité presque vides.

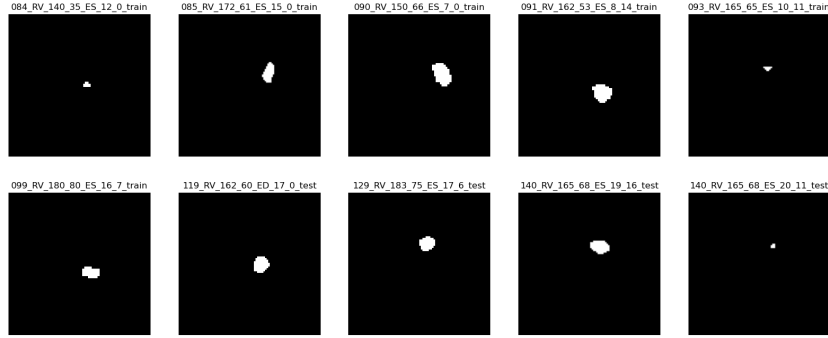


FIGURE 3.5 – Coupes non-vides parmi les outliers

Dans la suite, ces outliers sont écartés car ils n’expriment que peu d’information et nuisent à la clarté visuelle des UMAPs.

3.2.3 Position dans l’espace latent des points associés à des hauteurs différentes

Nous nous sommes ensuite intéressés à la répartition des codes associées à des coupes de hauteurs différentes. La hauteur est mesurée proportionnellement à la taille totale du coeur. Ainsi, la coupe la plus basse aura une hauteur égale à 0 tandis que la coupe la plus haute

aura une hauteur égale à 1. Les résultats sont présentés en figure 3.6. Les coupes vides sont omises dans la projection. On remarque que les coupes de hauteurs différentes sont encodées à des endroits distincts dans l'espace latent.

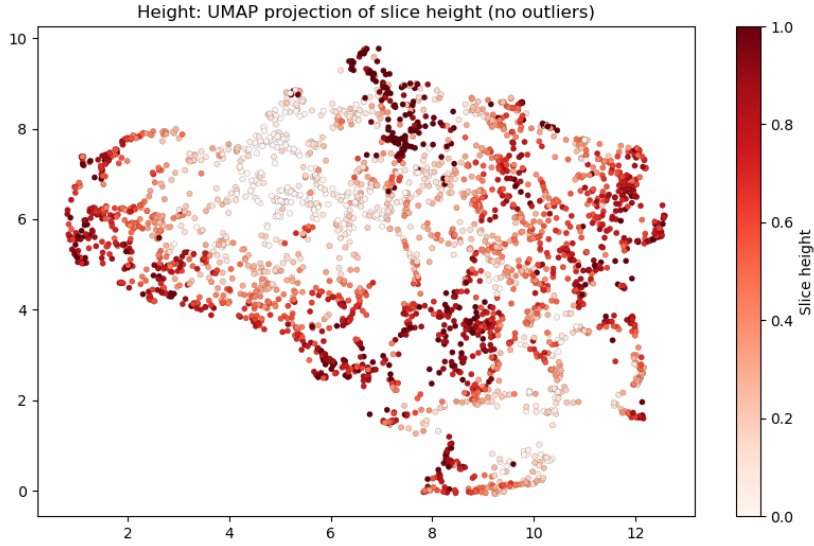


FIGURE 3.6 – UMAP des points associés à différentes hauteurs de coupe

3.2.4 Position dans l'espace latent des points associés à des pathologies différentes

Nous avons ensuite étudié la distribution dans l'espace latent des groupes de pathologie. Nous avons d'abord observé la répartition de tous les points, présentés en figure 3.7. Une nouvelle projection a été calculée en ne considérant que les coupes centrales positionnées à des hauteurs comprises entre 0.5 et 0.7. On observe des regroupements des points différents selon les pathologies. Le groupe des patients sains est projeté au centre de la UMAP. On observe un alignement des groupes de pathologie RV (ventricule droit anormal), groupe sain, DCM (ventricule gauche trop gros) et MINF. Cet axe traduit les anomalies sur les tailles des ventricules. D'un côté, les anomalies sont concentrées sur les ventricules gauches, de l'autre, les anomalies sont concentrées sur les ventricules droits, au centre, les patients sont sains. Le groupe HCM est distincts des autres et concerne les anomalies du myocarde.

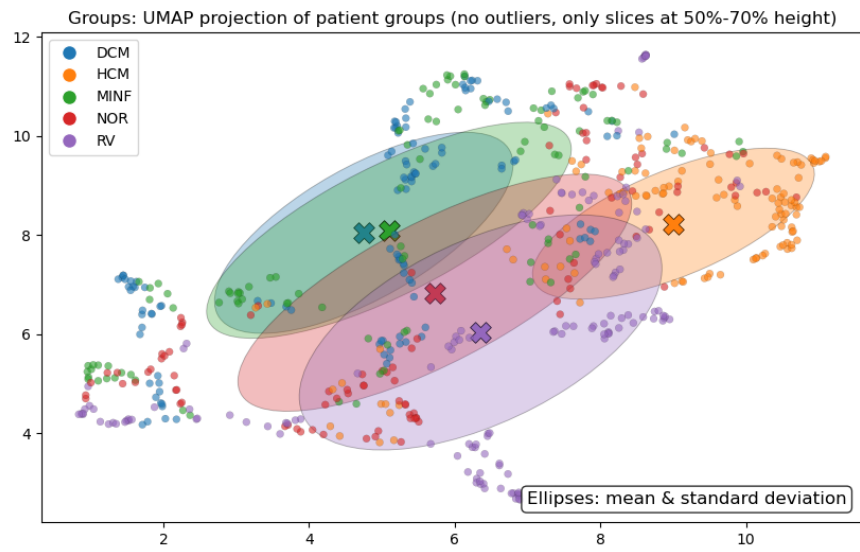


FIGURE 3.7 – Projection des groupes de patients (sans outliers, coupes entre 50 et 70% de hauteur).

Chapitre 4

Conclusion

Les résultats des parties 2.1.2 et 2.2.2 suggèrent que notre β -VAe est à la fois capable de reconstruire fidèlement des images d'IRM cardiaques et d'en générer de nouvelles. La recherche des paramètres optimaux 2.3 nous a permis d'obtenir un bon compromis entre les deux. L'analyse de l'espace latent permet de comprendre comment notre modèle encode les différentes propriétés des structures anatomiques.

Dans le papier d'origine [3], les chercheurs utilisent plusieurs critères pour quantifier la cohérence anatomique d'une IRM, pour se faire ils calculent par exemple s'il y a des trous dans les structures ou bien si certaines structures apparaissent plusieurs fois dans l'image. Ils utilisent un unique neurone entraîné en même temps que l'encodeur et le décodeur pour apprendre si une image respecte ou non ces critères structurants. La valeur en sortie de ce neurone est introduite dans la fonction de coût ce qui pousse le modèle à maximiser la cohérence des images générées.

Pour pouvoir mieux analyser l'espace latent, il serait pertinent d'entraîner le modèle en enlevant les slices vides (qui correspondent souvent à la première et la dernière slice d'un patient). En effet, l'encodage dans l'espace latent des images vides prend de la place dans la distribution de la gaussienne centrée réduite, ce qui pousse les images non vides à prendre moins de place dans l'espace latent. De plus, les images vides et non vides étant très différentes, il est nécessaire d'avoir une distance élevée entre les deux (cf Figure 3.3), ce qui fait que lorsque l'on génère des nouvelles images, les images vides sont surreprésentées. En les enlevant lors de l'entraînement, il y aurait plus de place dans l'espace latent pour les images ce qui permettrait de mieux pouvoir analyser la project UMAP et l'influence des différentes coordonnées de l'espace latent sur les structures des images.

Bibliographie

- [1] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis : Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11) :2514–2525, Nov 2018.
- [2] Carlo Biffi, Ozan Oktay, Giacomo Tarroni, Wenjia Bai, Antonio De Marvao, Georgia Doumou, Martin Rajchl, Reem Bedair, Sanjay Prasad, Stuart Cook, Declan O'Regan, and Daniel Rueckert. Learning interpretable anatomical features through deep generative models : Application to cardiac remodeling. in international conference on medical image computing and computer-assisted intervention (pp. 464-471). *Chemical Engineering Science*, 61(16) :5204–5220, 2006.
- [3] Nathan Painchaud, Youssef Skandarani, Thierry Judge, Olivier Bernard, Alain Lalande, and Pierre-Marc Jodoin. Cardiac mri segmentation with strong anatomical guarantees, 07 2019.