

Modèles de diffusion appliqués à l'imagerie médicale : augmentation de données et reconstruction d'images

Culture Sciences
de l'Ingénieur

La Revue
3E.I

Noah TOURNIER^{1,2} - Gatien SÉGUY^{1,3} - Alexandre NOEL^{1,4}

Édité le
15/10/2026

école —
normale —
supérieure —
paris—saclay —

¹ ENS Paris-Saclay, ² Centre Borelli, ³ Laboratoire SATIE, ⁴ CMAP

Cette ressource est issue des travaux de Noah Tournier dans le cadre du projet ARIA au Centre Borelli, de ceux de Gatien Séguéy menés lors de son stage de Master 2 FESup au laboratoire des systèmes et applications des technologies de l'information et de l'énergie (CNRS), ainsi que de ceux d'Alexandre Noël lors de son stage de Master 2 MVA au Centre de Mathématiques Appliquées de l'École Polytechnique.

Cette ressource fait partie du N° 121 de La Revue 3EI du 4^{ème} trimestre 2026.

L'imagerie médicale accompagne le patient à chaque étape du soin : dépistage, diagnostic et suivi d'un traitement. Elle repose sur plusieurs techniques : radiographie, scanner, IRM (imagerie par résonance magnétique), échographie ou encore tomographie par émission de positons. Dans beaucoup de ces techniques un appareil mesure d'abord un signal, puis une reconstruction de l'image est effectuée à partir de ces mesures. Par exemple, en IRM, l'appareil mesure ce qu'on appelle des coefficients de Fourier de l'image, en scanner, il mesure des projections de rayons X sous de nombreux angles.

Cette chaîne d'acquisition pose un premier problème. Afin d'avoir des images de meilleure qualité il est nécessaire d'effectuer plus de mesure ce qui augmente la durée de l'examen et son coût. Pour pallier ce problème il faut donc réduire le nombre de mesures ou recourir à un équipement moins complexe. L'image calculée est alors dégradée, elle est bruitée, floue ou marquée d'artefacts. Le problème devient alors un problème de reconstruction d'images : comment retrouver une image exploitable malgré des mesures incomplètes ?

Un second problème concerne l'analyse automatique des images. Des réseaux de neurones aident de plus en plus au diagnostic, par exemple pour repérer une tumeur. Pour ce faire, ils sont entraînés sur des jeux de données comprenant énormément d'images. Or un jeu de données médicales est difficile à constituer, il faut acquérir les examens, obtenir les autorisations et anonymiser les données. Il faut aussi que des médecins annotent les images. De plus, les images proviennent souvent d'appareils et de protocoles différents, ce qui peut biaiser l'apprentissage. Enfin, les maladies rares sont peu représentées. Le réseau apprend alors surtout à reconnaître les cas fréquents, et il est moins performant sur les cas rares, qui sont pourtant parmi les plus importants à détecter. Une solution est de compléter le jeu de données avec de nouvelles images, sans nouveaux examens ni nouveaux patients : c'est l'augmentation de données.

L'intelligence artificielle générative apporte une réponse à ces deux problèmes. Un modèle génératif apprend, à partir d'exemples, ce qu'est une image médicale plausible. Il peut ensuite créer de nouvelles images du même type. Parmi ces modèles, les modèles de diffusion occupent une place importante depuis 2020. Leur idée est simple : on apprend à un réseau de neurones à

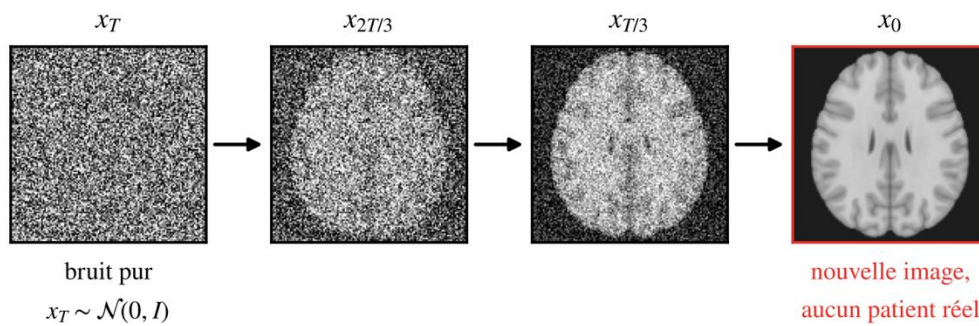
retirer du bruit d'une image, un peu à la fois. En partant d'un bruit pur et en le retirant étape par étape, on voit apparaître une image. Ces modèles produisent des images variées et de bonne qualité. On peut aussi les guider par une information connue, comme une classe (bénin ou malin), un masque ou une mesure.

Cet article présente ces modèles et deux de leurs usages en imagerie médicale :

- **L'augmentation de données** : créer de nouvelles images pour mieux représenter les cas rares ;
- **La reconstruction d'images** : retrouver une image exploitable à partir d'une acquisition bruitée, floue ou incomplète.

Ces deux usages reposent sur la même connaissance : le modèle sait ce qu'est une image médicale plausible. Cependant ils sont utilisés dans des cas bien différents. Une image générée ne représente aucun patient précis alors qu'une image reconstruite représente un patient réel et doit rester fidèle à ses mesures. Un modèle qui ajoute un détail plausible mais faux met alors le diagnostic en danger. On parle alors d'hallucination.

Application 1 : générer de nouvelles images (augmentation de données)



Application 2 : reconstruire l'image d'un patient

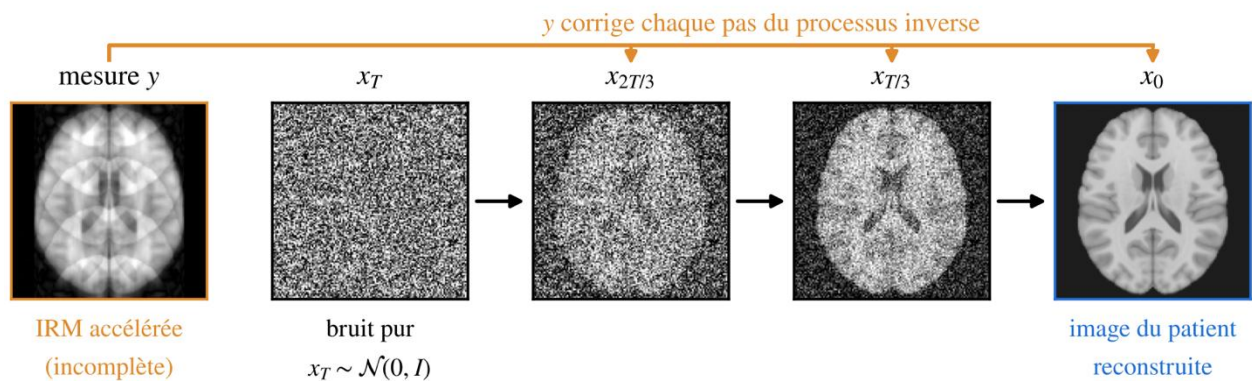


Figure 1 : Les deux usages étudiés. En reconstruction, le modèle doit respecter la mesure du patient.

1 - Modèles de diffusion : fondements théoriques

Un modèle génératif apprend, à partir d'un jeu d'exemples, à produire de nouvelles données. Dans notre cas ce sont des images médicales. Ces images ne sont pas des copies de celles du jeu initial, cependant elles conservent les propriétés générales, telles que les contrastes, les structures anatomiques ou les textures caractéristiques du type d'examen considéré.

Les premiers modèles de diffusion modernes apparaissent en 2015 [1]. Les DDPM (Denoising Diffusion Probabilistic Models) popularisent ensuite la méthode en 2020 [2]. Leur apprentissage est

stable et ils permettent de contrôler la génération. Ces deux qualités expliquent leur intérêt en imagerie médicale.

1.1 - Principe général

Un modèle de diffusion utilise deux processus opposés :

- Le processus direct qui ajoute du bruit à une image et la déstructure jusqu'à ce qu'elle ne soit plus qu'un bruit blanc gaussien. Il est fixé à l'avance.
- Le processus inverse qui retire ce bruit pas à pas et permet de récupérer l'image originale. Il est appris par un réseau de neurones

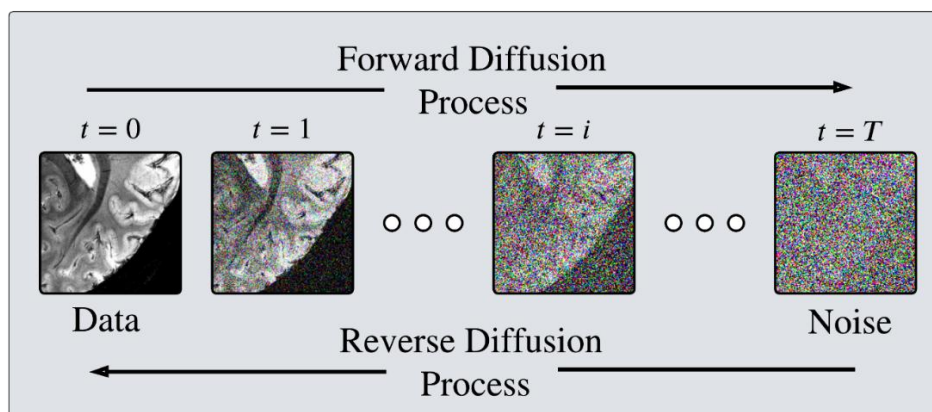


Figure 2 : Processus direct et inverse d'un modèle de diffusion, illustrés sur un patch d'une coupe IRM axiale de cerveau. Le processus direct ajoute progressivement du bruit gaussien jusqu'à un bruit pur. Le processus inverse, appris par le réseau, reconstruit l'image en retirant le bruit étape par étape.

Une fois l'apprentissage terminé, le processus de génération s'effectue de la manière suivante : une image ne contenant que du bruit est créée, le réseau applique alors le processus inverse sur ce bruit, et une nouvelle image est générée étape par étape.

Le mot *débruiter* peut prêter à confusion. En effet, aucune image n'est cachée dans le bruit de départ, le réseau construit une nouvelle image, en la rapprochant à chaque étape des images apprises. Deux bruits de départ différents donnent donc deux images différentes.

Une expérience de pensée :

Prenons un jeu de coupes IRM du cerveau. Le modèle repère des propriétés communes : la forme du crâne, les contrastes entre les tissus, certaines symétries, la forme des sillons, etc... Il apprend aussi les variations normales entre les patients. Il peut ensuite créer une nouvelle coupe qui respecte ces propriétés.

1.2 - Le processus direct

On note x_0 l'image de départ. Le processus direct produit les images $x_1, x_2, \dots, x_T$. L'image est légèrement bruitée et déstructurée à chaque étape pour qu'à la fin, x_T approxime un bruit blanc gaussien. Par déstructuration, on entend la disparition progressive de l'organisation spatiale de l'image : les contours, les formes et les contrastes s'effacent peu à peu sous l'effet du bruit ajouté.

Il n'est pas nécessaire de calculer toutes les étapes. On peut obtenir directement x_t à partir de x_0 :

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I) \quad (1)$$

Le terme ε est un bruit gaussien de même taille que l'image et le coefficient $\bar{\alpha}_t$ décroît de 1 vers 0. L'équation contient donc deux parties :

- $\sqrt{\bar{\alpha}_t} x_0$ qui est la part de signal encore conservée ;
- $\sqrt{1 - \bar{\alpha}_t} \varepsilon$ qui est la part de bruit introduite.

Les racines carrées s'appliquent aux amplitudes. Les coefficients $\bar{\alpha}_t$ et $1 - \bar{\alpha}_t$ donnent alors les parts de puissance. Le rapport signal sur bruit vaut :

$$RSB(t) = \frac{\bar{\alpha}_t}{1 - \bar{\alpha}_t} \quad (2)$$

Au début du processus, $\bar{\alpha}_t$ est proche de 1 et l'image reste visible, tandis qu'à la fin, $\bar{\alpha}_t$ tend vers 0 et le bruit domine entièrement l'image. La suite des valeurs de $\bar{\alpha}_t$ forme ce que l'on appelle le *planning de bruitage*. Celui-ci est fixé avant l'apprentissage, car il ne dépend pas des données : il suffit de définir à l'avance la quantité de bruit à ajouter à chaque étape, le réseau n'a donc pas à l'apprendre.

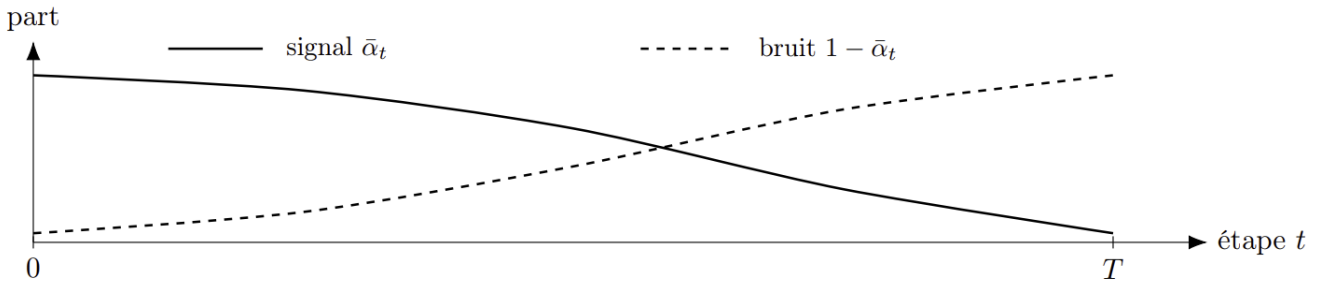


Figure 3 : Lecture qualitative du planning de bruitage : la part de signal décroît tandis que la part de bruit augmente. Les formes exactes dépendent du planning choisi.

1.3 - Le processus inverse et l'apprentissage

Le réseau de neurones est noté ε_θ . Il s'agit d'une fonction comportant un grand nombre de paramètres, notés θ , dont les valeurs sont modifiées pendant l'apprentissage. Il reçoit deux informations : l'image bruitée x_t et l'étape t . Son but est d'estimer le bruit présent dans x_t . Cette estimation permet ainsi de calculer une image un peu moins bruitée.

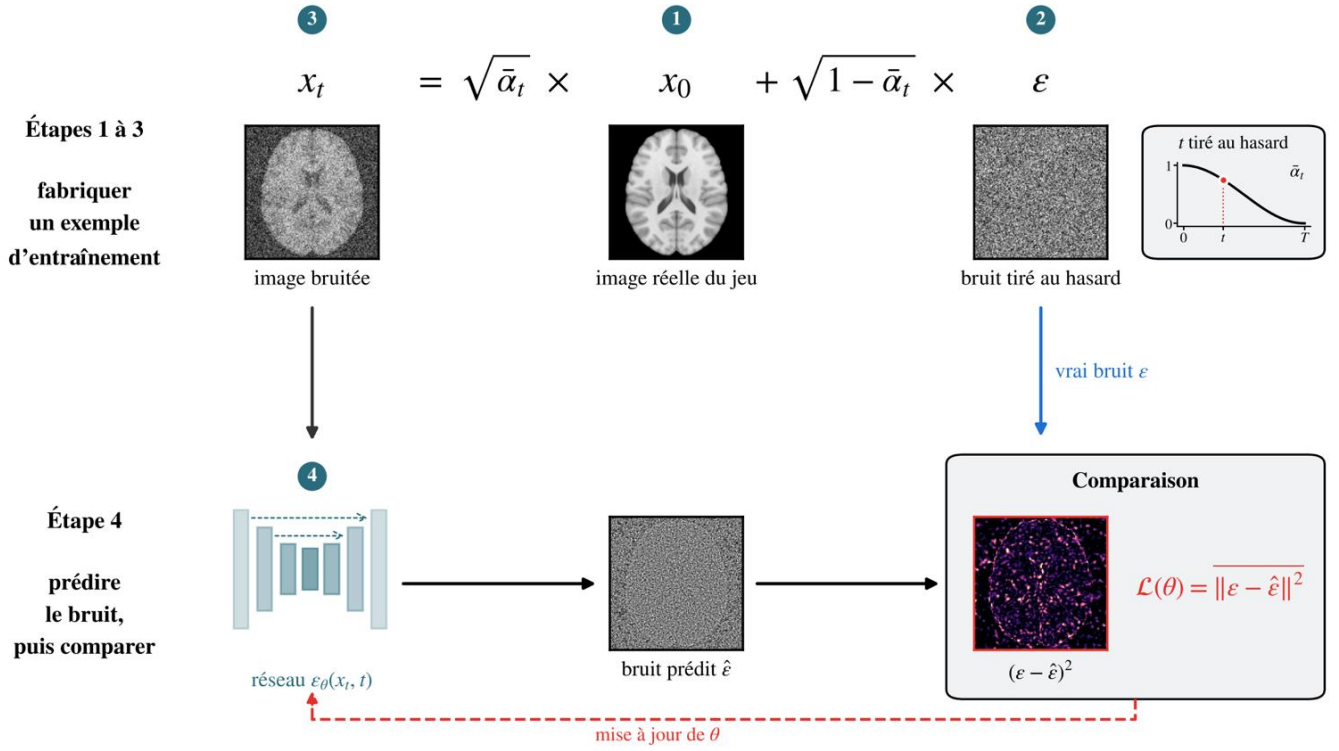
L'apprentissage suit quatre étapes simples :

1. On choisit une image x_0 .
2. On choisit une étape t et un bruit ε .
3. On calcule l'image bruitée x_t .
4. Le réseau prédit le bruit. On compare alors sa réponse au bruit réel ε .

La fonction de coût est une erreur quadratique moyenne entre le bruit réel et le bruit prédit :

$$\mathcal{L}(\theta) = \mathbb{E}[\|\varepsilon - \varepsilon_\theta(x_t, t)\|^2] \quad (3)$$

Le bruit ε est connu, car le programme vient de le générer, il sert de réponse attendue pendant l'apprentissage. La norme mesure l'erreur sur tous les pixels et l'espérance représente la moyenne sur les images, les étapes et les bruits. L'apprentissage règle les paramètres θ pour réduire cette erreur.



Cet apprentissage est un problème de régression, il ne met pas deux réseaux en concurrence, comme dans un GAN (Generative Adversarial Network). Il est donc souvent plus stable.

Le réseau est le plus souvent un U-Net, conçu à l'origine pour la segmentation d'images médicales, c'est-à-dire pour isoler un organe ou une lésion. Ce type de réseau est composé de deux branches, une branche descendante et une branche montante. La branche descendante analyse l'image à plusieurs échelles et la branche montante retrouve progressivement la résolution initiale, des connexions directes entre les deux permettent de conserver les détails fins. Le réseau reçoit aussi l'étape t en entrée, ce qui lui permet d'adapter son traitement au niveau de bruit présent dans l'image.

Pour générer une image, ce réseau est appliqué de façon répétée : il transforme x_T en x_{T-1} , puis x_{T-1} en x_{T-2} , et ainsi de suite jusqu'à obtenir x_0 . Cette répétition donne de bons résultats mais coûte cher en temps de calcul, c'est pourquoi deux familles de méthodes cherchent à limiter ce coût. Les DDIM (Denoising Diffusion Implicit Models) qui permet de réduire le nombre d'étapes nécessaires, et les modèles de diffusion latente qui travaillent sur une version compressée de l'image.

1.4 - Génération conditionnelle et guidage

Sans information supplémentaire, le contenu de l'image générée est aléatoire : le modèle peut produire un cerveau sain comme un cerveau malade. Pour compléter un jeu de données, on veut au contraire choisir ce que l'on génère, par exemple une lésion rare. On fournit alors au réseau une information supplémentaire c , appelée condition. Le réseau reçoit trois entrées :

$$\hat{\epsilon} = \epsilon_{\theta}(x_t, t, c) \quad (4)$$

La condition peut être une classe (« sain », « lésion »), un masque qui indique où placer la lésion, ou une autre image du même patient. Elle ne fixe pas tous les pixels, elle indique seulement ce que l'image doit respecter. Par exemple, la condition « lésion dans le lobe frontal » autorise encore

plusieurs formes et plusieurs anatomies. À partir du même bruit de départ, chaque condition conduit ainsi à une image différente (figure 5).

En pratique, on règle aussi la force avec laquelle la condition est imposée, à l'aide d'un paramètre appelé force du guidage [6]. Une force trop faible respecte mal la condition. Une force trop grande impose bien la condition, mais réduit la diversité des images et peut créer des défauts. Il faut donc trouver un compromis. Nous retrouverons ce type de réglage dans la seconde application, la reconstruction d'images.

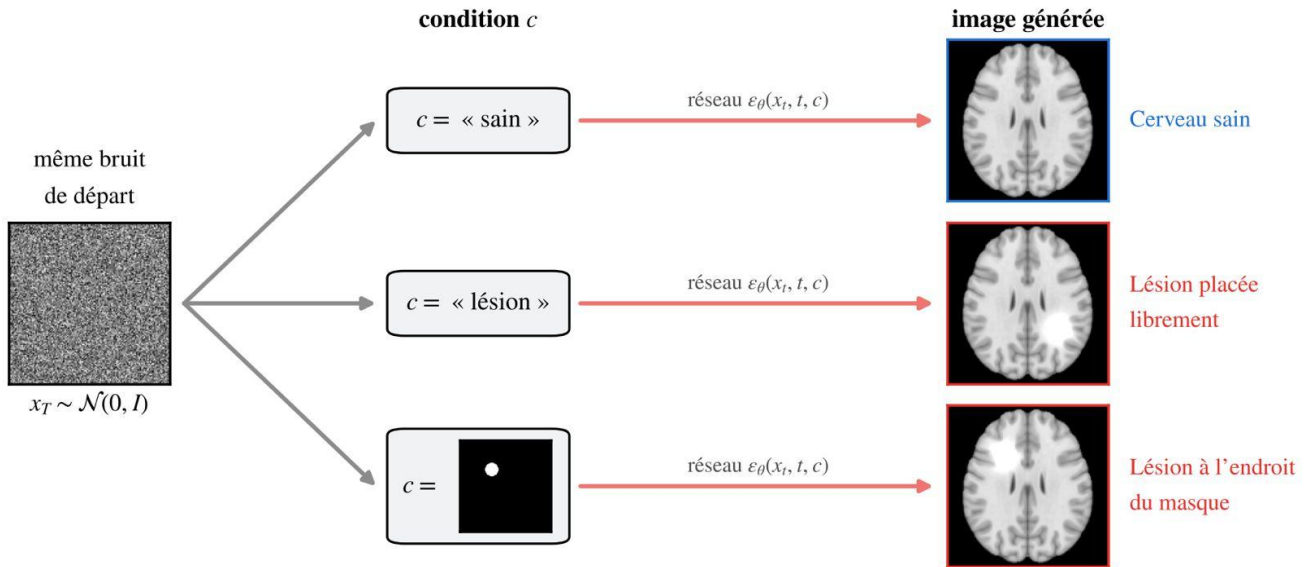


Figure 5 : À partir du même bruit de départ, la condition c oriente la génération. Images illustratives : la lésion est simulée.

1.5 - Ce que le modèle apprend : le score

Jusqu'ici, nous avons présenté le réseau comme une fonction qui prédit le bruit présent dans une image. La même information peut être interprétée autrement.

Ainsi, le score donne une autre lecture du modèle. Il s'écrit :

$$s_{\theta}(x_t, t) \simeq \nabla_{x_t} \log p_t(x_t). \quad (5)$$

Ce gradient indique dans quel sens modifier les pixels. La modification rend l'image plus probable selon le modèle. Pour un bruit gaussien, le score est directement lié au bruit prédit :

$$s_{\theta}(x_t, t) = -\frac{\varepsilon_{\theta}(x_t, t)}{\sqrt{1 - \bar{\alpha}_t}} \quad (6)$$

Le score pousse donc l'image vers une forme plausible. Lors d'une reconstruction, un autre terme la pousse vers la mesure du patient.

À retenir avant les applications

Le modèle de diffusion apprend progressivement à transformer une image bruitée en une image cohérente et réaliste. Lors de cette génération, une condition, comme une description textuelle ou une image de référence, peut guider le modèle. Elle influence ainsi la manière dont l'image est reconstruite et permet de mieux contrôler le contenu de l'image finale.

1.6 - Approche voisine : les normalizing flows

Les modèles de diffusion ne sont pas la seule façon de transformer un bruit simple en une image complexe. Une autre famille de modèles, les normalizing flows, permet également de générer de nouvelles images.

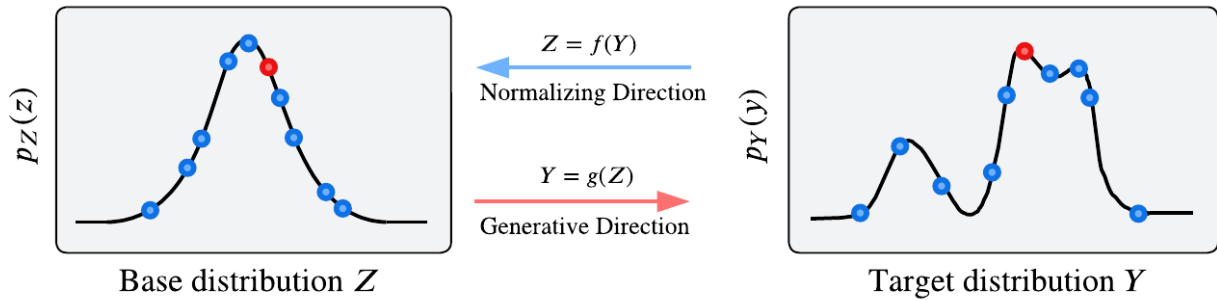


Figure 6 : Principe d'un normalizing flow : une transformation inversible g relie une distribution simple Z à la distribution complexe des images Y .

Au lieu d'ajouter du bruit étape par étape, un normalizing flow apprend une seule fonction g , inversible, qui transforme directement une variable simple z (par exemple gaussienne) en une image y :

$$y = g(z).$$

Pour générer une nouvelle image, il suffit de tirer un z au hasard (le point rouge sur le schéma de la Figure 6) et de calculer $y = g(z)$. Une seule opération, sans les nombreuses étapes de débruitage utilisées en diffusion.

La difficulté est que g doit aussi permettre de calculer la probabilité d'une image réelle y , pour pouvoir entraîner le modèle. En notant $f = g^{-1}$, cette probabilité s'écrit :

$$p_Y(y) = p_Z(f(y)) |\det \nabla f(y)|.$$

Le premier terme mesure si $z = f(y)$ est une valeur plausible pour la loi simple p_Z . Le second terme, le déterminant du jacobien, corrige un effet de changement d'échelle. En effet, transformer l'espace peut « étirer » ou « comprimer » les probabilités, un peu comme étirer un segment de longueur 1 en un segment de longueur 2 divise par deux la densité de probabilité par unité de longueur. Ce terme généralise cette idée à une image entière.

L'entraînement consiste à choisir les paramètres de g de façon à rendre $p_Y(y_i)$ la plus grande possible pour chaque image réelle y_i du jeu de données.

Il est important de noter que les modèles de diffusion dominent largement le domaine de la génération d'images, de molécules, de protéines etc... Et ce parce que ces modèles permettent d'obtenir des résultats de haute qualité avec une bonne variation. Les normalizing flows classiques, comme ceux présentés plus haut, restent moins utilisés à grande échelle, leur principal problème étant la génération d'images complexes.

Une évolution récente change cependant la donne : le *flow matching*. Cette méthode reprend l'idée d'une transformation continue déjà présente dans les normalizing flows, mais l'entraîne d'une façon proche de celle des modèles de diffusion, sans passer par un bruitage étape par étape. Le résultat est une génération nettement plus rapide que la diffusion classique, pour une qualité comparable. Cette approche commence à être utilisée dans des systèmes de génération d'images, de conception de molécules, en complément ou en remplacement de la diffusion traditionnelle.

Sur des images médicales comme les mammographies, la diffusion a tendance à mieux restituer la texture des tissus, au prix d'un temps de génération plus long. Les normalizing flows restent une alternative intéressante lorsque la rapidité de génération est prioritaire.

2 - Application 1 : augmenter un jeu de données médicales

2.1 - Pourquoi l'augmentation classique ne suffit pas toujours

Face au manque de données évoqué en introduction, la première solution qui vient à l'esprit consiste à multiplier artificiellement les images existantes. C'est le principe de l'augmentation classique, une image du jeu de données est tournée, recadrée, ou voit son contraste légèrement modifié, afin de produire une variante qui viendra enrichir l'entraînement. Ces transformations ont l'avantage d'être simples à mettre en œuvre et très rapides à calculer. Elles présentent toutefois une limite fondamentale : elles ne font que modifier la présentation d'une image déjà existante, sans jamais en créer une nouvelle. Une lésion que l'on fait pivoter reste, anatomiquement, exactement la même lésion. Le jeu de données s'agrandit donc en apparence, mais la diversité réelle des cas qu'il représente reste très limitée.

C'est précisément là qu'un modèle génératif apporte quelque chose. Plutôt que de transformer une image existante, il apprend les caractéristiques générales d'un ensemble d'images pour en produire de nouvelles, qui n'ont pas d'équivalent direct dans le jeu d'entraînement. Il devient alors possible de générer une anatomie différente, ou une lésion présentant une forme ou une localisation inédite. Cette capacité ne rend pas pour autant l'augmentation classique obsolète : les deux approches poursuivent des objectifs différents et se complètent naturellement, la première apportant du volume à moindre coût, la seconde une véritable diversité de contenu.

L'augmentation de données par diffusion suit un enchaînement de plusieurs étapes. Le modèle est d'abord entraîné à partir des images réelles. Une fois cet apprentissage terminé, il est utilisé pour produire de nouvelles images synthétiques. Ces images ne sont pas directement intégrées au jeu de données. En effet, elles passent d'abord par un contrôle qualité, destiné à écarter les doublons ainsi que les cas incohérents. Les images jugées satisfaisantes viennent alors enrichir le jeu de données initial. Ce jeu enrichi peut enfin être utilisé pour entraîner ou améliorer un modèle sur une tâche médicale donnée, par exemple classification ou détection d'anomalies.

2.2 - Choisir la condition

La condition doit correspondre au besoin médical et dépend des annotations disponibles.

- Une classe (« malin », « bénin », ...) peut représenter une maladie. Elle contrôle peu la forme et la position de la lésion.
- Un masque fixe la forme et la position d'une région. Il fournit aussi une annotation pour la future image.
- Une autre modalité peut servir d'entrée. Par exemple, une séquence IRM peut guider la création d'une autre séquence.
- Des métadonnées peuvent préciser l'âge ou le protocole d'acquisition. Elles peuvent aussi introduire des biais.

Garder les proportions du jeu de données

Lorsqu'une image est découpée en patches, la proportion de patches contenant une lésion doit rester proche de la proportion de surface occupée par cette lésion dans l'image complète :

$$p_{\text{lésion}} \approx \frac{A_{\text{lésion}}}{A_{\text{image}}}$$

Si cette proportion n'est pas respectée, le modèle peut apprendre à générer presque uniquement des images sans lésion, ce qui est irréaliste dans un contexte de dépistage.

Le jeu d'entraînement reste essentiel. Le modèle ne peut pas apprendre ce qu'il ne voit pas. Une population absente du jeu restera absente des images générées.

2.3 - Entraînement et production des images synthétiques

Les images doivent avoir une orientation, une résolution et une échelle d'intensité cohérentes. Un recalage peut aussi être nécessaire. Ces traitements ne doivent pas supprimer une information clinique.

Un exemple concret illustre ce compromis. Sur des mammographies découpées en patches, un patch dont la taille est trop faible par rapport à celle d'une lésion risque de ne contenir qu'une quantité limitée d'informations pertinentes. Le réseau ne verra alors que des zones de tissu presque uniformes, ce qui augmentera le risque d'apprendre à reproduire une intensité de gris moyenne plutôt qu'une texture réaliste. Le choix de la taille des patches constitue donc un paramètre important, puisqu'il conditionne la quantité de contexte disponible et permet au modèle d'apprendre des structures médicales plausibles.

Le partage des données est effectué au niveau du patient. En effet, les images provenant d'un même patient sont systématiquement conservées dans le même ensemble de données. Cela permet d'éviter qu'un patient soit représenté à la fois dans les données d'apprentissage et dans celles de test, ce qui pourrait entraîner une fuite de données, notamment lorsque plusieurs coupes très proches sont disponibles.

Une fois le modèle entraîné, plusieurs images sont générées afin d'augmenter la représentation des classes les plus rares. La force du guidage influence le respect des conditions imposées lors de la génération, tandis que le nombre d'étapes de diffusion influe principalement sur le temps de calcul et la qualité des images produites.

Les images générées doivent ensuite faire l'objet d'un contrôle qualité. Les doublons ainsi que les images présentant des incohérences sont éliminés. Enfin, un échantillon représentatif des images générées doit être examiné par un expert médical afin d'évaluer leur plausibilité et leur cohérence avec les données cliniques attendues.

2.4 - Comment évaluer l'utilité des images générées

Une image réaliste n'est pas toujours utile. Il faut donc vérifier plusieurs points.

Question	Évaluation possible	Limite à garder en tête
Réalisme	Lecture en aveugle, recherche d'artefacts, distributions d'intensité	Un aspect réaliste ne garantit pas la justesse clinique

Question	Évaluation possible	Limite à garder en tête
Fidélité à la condition	Classifieur externe, recouvrement avec un masque, lecture experte	Un outil automatique peut partager les biais du générateur
Diversité	Distances entre échantillons, couverture des sous-groupes, détection de doublons	Une forte dispersion peut aussi provenir d'artefacts
Non-mémorisation	Recherche des plus proches voisins dans le jeu d'entraînement	Une distance globale peut manquer une copie locale
Utilité	Performance du modèle aval sur un test composé uniquement de données réelles	C'est le critère décisif mais il dépend de la tâche choisie

Table 1 : Une évaluation multidimensionnelle des données synthétiques.

Pour ce qui est du réalisme il est aussi possible de soumettre les images à un test visuel : un radiologue doit deviner, pour chaque image, si elle est réelle ou générée. Un taux de réussite proche de 50 % indique que les images sont difficiles à distinguer. Ce type de vérification bien que le plus fiable est aussi le plus long à organiser.

La distance de Fréchet (FID) est souvent utilisée puisqu'elle permet de comparer la distribution des images générées à celle des images réelles dans un espace de caractéristiques appris. Une valeur faible indique une bonne ressemblance globale, mais une valeur proche de zéro peut aussi signaler un manque de diversité. On peut aussi vérifier que toutes les catégories de patients (âge, sexe, type d'appareil) sont bien représentées parmi les images générées, et pas seulement les cas les plus fréquents du jeu d'entraînement.

Enfin, il est possible d'entraîner la même architecture de réseau avec plusieurs jeux de données : données réelles seules, augmentation classique, puis ajout d'images synthétiques. Le jeu de test reste réel et indépendant. On compare ensuite la sensibilité, la spécificité et les résultats pour chaque sous-groupe.

Exemple de protocole pour une classe rare

Un jeu contient 2 000 examens sans anomalie et 150 examens avec une lésion rare. Le modèle crée de nouveaux cas de cette lésion. On entraîne alors trois classifieurs : (A) avec les données réelles ; (B) avec une augmentation classique ; (C) avec les données réelles et synthétiques. On peut alors comparer les trois classifieurs sur un même jeu de données test afin de juger la qualité des images générées.

2.5 - Questions de confidentialité

Une image synthétique ne correspond pas directement à un examen médical réel et peut ainsi faciliter le partage des données. Toutefois, sa génération ne garantit pas nécessairement l'anonymat des patients. En effet, un modèle surappris peut être susceptible de reproduire, de manière partielle ou quasi identique, certaines images présentes dans les données d'entraînement. Il est donc nécessaire de rechercher d'éventuelles copies ou similarités excessives et d'évaluer le risque d'identification des patients. Enfin, la traçabilité et la provenance des données utilisées pour entraîner le modèle doivent être maintenues.

3 - Application 2 : Reconstruire une image à partir de mesures

Dans la première application, le modèle crée des images qui ne correspondent à aucun patient. Ici, la situation est différente. L'image recherchée est celle d'un patient réel, et l'appareil en a fourni une mesure incomplète ou dégradée. Le modèle de diffusion ne sert plus à inventer une image, mais à compléter l'information qui manque dans la mesure, sans jamais la contredire.

3.1 - Du dispositif d'acquisition au problème inverse

Un dispositif d'imagerie ne mesure pas nécessairement l'image directement. Il mesure un signal physique dépendant des propriétés de l'objet observé. Ce processus peut être décrit à l'aide d'un modèle physique, qui s'écrit :

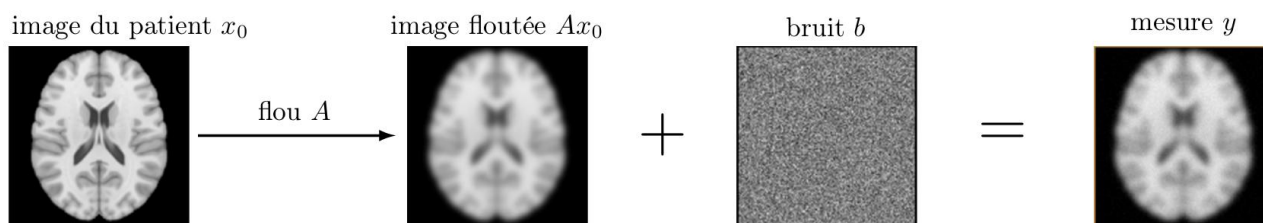
$$y = Ax_0 + b \quad (7)$$

Dans cette équation, x_0 est l'image recherchée, y est la mesure disponible, A représente l'appareil et b représente le bruit de mesure.

En IRM, y contient les coefficients de Fourier mesurés. En tomographie, il correspond aux projections acquises, tandis que, dans le cas d'un problème de défloutage, A représente l'opérateur de flou.

La mesure y étant connue, l'objectif est de retrouver l'image originale x_0 . Il s'agit d'un problème inverse. Sa résolution n'est pas nécessairement unique : plusieurs images peuvent être compatibles avec une même mesure, ou avec des mesures très proches. Il est donc nécessaire d'introduire des informations supplémentaires sur les images recherchées. Ces informations sont appelées a priori (ou *prior*).

Problème direct : $y = Ax_0 + b$



Problème inverse : retrouver x_0 à partir de y

Figure 7 : Le problème direct va de l'image vers la mesure : l'appareil floute l'image, puis le bruit s'ajoute. La reconstruction suit le chemin inverse, beaucoup plus difficile.

Une reconstruction classique minimise par exemple :

$$\hat{x} = \arg \min_x \underbrace{\|y - Ax\|^2}_{\text{fidélité aux mesures}} + \underbrace{\lambda R(x)}_{\text{régularisation}} \quad (8)$$

Le premier terme mesure l'accord avec les données (l'image dégradée), il doit être faible lorsque l'image explique bien la mesure. Le second terme impose une propriété choisie (un prior), il peut, par exemple, favoriser une image lisse. Enfin, le coefficient λ règle l'équilibre entre les deux termes.

Un modèle de diffusion apporte un prior appris de meilleure qualité. Il remplace alors une partie des règles choisies à la main.

3.2 - Lecture bayésienne : combiner mesure et connaissance a priori

La règle de Bayes exprime la même idée :

$$p(x | y) = \frac{p(y | x) p(x)}{p(y)} \propto p(y | x) p(x) \quad (9)$$

Le terme $p(x)$ représente la plausibilité de l'image x . Le terme $p(y | x)$, appelé vraisemblance, indique dans quelle mesure l'image x est compatible avec la mesure observée y . Enfin, $p(x | y)$ combine ces deux informations.

Pour un bruit gaussien de variance σ_b^2 , la vraisemblance s'écrit :

$$p(y | x) \propto \exp\left(-\frac{\|y - Ax\|^2}{2\sigma_b^2}\right) \quad (10)$$

Une image qui explique mal la mesure observée a une faible vraisemblance. En effet, lorsque le niveau de bruit est élevé, plusieurs images peuvent être compatibles avec les données mesurées. À l'inverse, lorsque la mesure est plus précise, elle contraint davantage la reconstruction et réduit l'éventail des solutions possibles.

Deux forces pendant la reconstruction.

Le modèle de diffusion rend l'image plausible et la vraisemblance permet d'imposer la mesure du patient. C'est pourquoi les deux termes sont indispensables. Le prior, utilisé seul, peut conduire à une image visuellement réaliste, mais qui peut porter de fausses informations. À l'inverse, les mesures seules peuvent être insuffisantes pour reconstruire une image de bonne qualité.

Toute la difficulté consiste à trouver un équilibre entre ces deux termes. On peut assimiler ce compromis à un curseur permettant d'ajuster l'importance accordée aux connaissances acquises par le modèle sur les images (le *prior*), par rapport aux informations réellement fournies par les mesures du patient. Dans les méthodes de reconstruction classiques, le coefficient λ de l'équation (8) joue ce rôle. La figure 8 illustre l'influence de ce paramètre sur le résultat du défloutage présenté dans la figure 7.

Si l'on se fie trop à ce que l'appareil a mesuré, le modèle n'aide presque pas. La reconstruction cherche alors à expliquer chaque détail de la mesure, y compris son bruit. Ce bruit est fortement amplifié et l'image devient inutilisable : le problème n'est pas résolu.

Si l'on se fie trop à ce que le modèle a appris, l'image devient nette et propre. Mais elle ressemble davantage aux images vues pendant l'apprentissage qu'au patient. Le modèle invente des détails que la mesure ne contient pas : c'est l'hallucination évoquée en introduction. Sur la figure 8, les sillons de l'image reconstruite sont ceux d'un autre cerveau.

Entre ces deux extrêmes, un réglage donne l'image la plus fidèle au patient. Ce réglage n'est pas universel : il dépend de l'appareil, du niveau de bruit et de l'image elle-même.

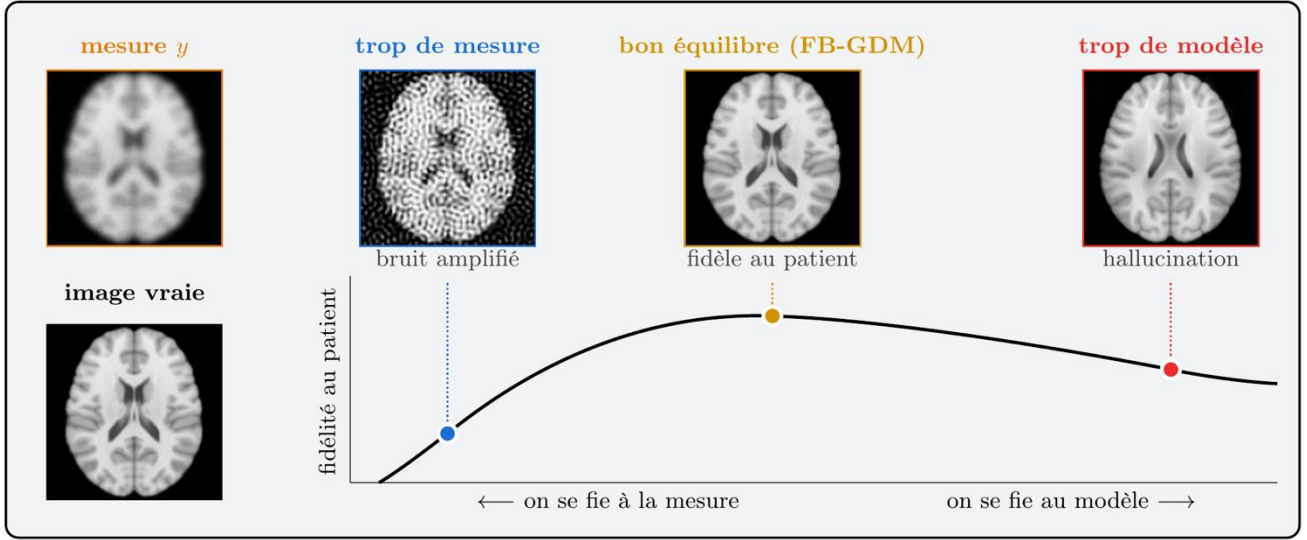


Figure 8 : Le compromis entre mesure et prior sur l'exemple de la figure 7.

3.3 - Guidage par les mesures

Revenons au processus inverse de la partie 1. Lors de la génération, chaque pas utilise le score $\nabla_{x_t} \log p_t(x_t)$ de la section 1.5. Ce score pousse l'image vers ce que le modèle a appris, c'est-à-dire vers une image plausible. Pour reconstruire, on veut davantage : une image plausible, mais aussi compatible avec la mesure y du patient. On remplace donc ce score par le score conditionnel $\nabla_{x_t} \log p_t(x_t | y)$, qui tient compte de la mesure. La règle de Bayes (9), appliquée à x_t , le sépare en deux termes :

$$\nabla_{x_t} \log p_t(x_t | y) = \underbrace{\nabla_{x_t} \log p_t(x_t)}_{\text{score à priori}} + \underbrace{\nabla_{x_t} \log p_t(y | x_t)}_{\text{terme de mesure}} \quad (11)$$

Le terme $p(y)$ de la règle de Bayes disparaît, car il ne dépend pas de x_t . Le premier terme est le score de la section 1.5 : il est fourni par le réseau grâce à l'équation (6). Il permet de « pousser » l'image vers ce que le modèle a appris. Le second terme « pousse » l'image vers ce que l'appareil a mesuré. On retrouve le même principe que celui des deux termes de la section 3.2.

Le processus inverse reste celui de la génération, seul le score change. En effet, à chaque pas, l'image est à la fois rendue plausible et rapprochée de la mesure. On dit que la mesure guide la diffusion (figure 1).

La mesure y dépend de l'image finale x_0 , et non de l'image bruitée x_t . Pour l'évaluer, le réseau prédit d'abord à quoi ressemblera l'image finale. On simule la mesure que donnerait cette image, puis on la compare à y . L'écart indique dans quel sens corriger l'image.

Ce calcul n'est qu'approché. En pratique, on pondère donc le terme de mesure par un coefficient η_t :

$$\nabla_{x_t} \log p_t(x_t | y) \approx \nabla_{x_t} \log p_t(x_t) + \eta_t \nabla_{x_t} \log p_t(y | x_t) \quad (12)$$

Le compromis de la section 3.2 apparaît ici directement. Si η_t est trop petit, ce que le modèle a appris l'emporte : l'image est plausible, mais elle peut contenir des hallucinations. Si η_t est trop grand, ce que l'appareil a mesuré l'emporte : le bruit de la mesure est amplifié. Le coefficient η_t est donc le curseur de la figure 8.

Deux méthodes de référence reposent sur ce principe : DPS (Diffusion Posterior Sampling) [9] et Π GDM (Pseudoinverse-Guided Diffusion Models) [10]. Ces méthodes donnent de très bons résultats, mais elles nécessitent toutes deux un réglage manuel de certains paramètres. Avant la reconstruction, l'utilisateur doit définir le poids accordé aux mesures ainsi que, dans certains cas, d'autres paramètres. De plus, Π GDM requiert une estimation du niveau de bruit associé aux mesures de l'appareil.

Comment choisir ces réglages ? En pratique, on essaie plusieurs valeurs et on garde celle qui donne l'image la plus proche de l'image réelle. Cette démarche fonctionne en laboratoire, sur des images simulées. Elle ne fonctionne pas à l'hôpital, car l'image réelle du patient est justement ce que l'on cherche. De plus, le bon réglage change avec l'appareil, le protocole et le niveau de bruit. Un réglage choisi pour un examen peut donner une image bruitée, ou une hallucination, sur l'examen suivant.

La méthode FB-GDM (Fully-Bayesian Guided Diffusion Models) [11] supprime ce réglage manuel. Développée par Gatien Séguy, l'un des auteurs du présent article, en collaboration avec Thomas Rodet, enseignant chercheur au laboratoire SATIE et DER SIEN de l'ENS Paris-Saclay, elle règle elle-même le bouton, à chaque pas, à partir de la mesure seule. L'idée est simple. À chaque pas, deux sources d'information sont disponibles : ce que le réseau prédit et ce que l'appareil a mesuré. FB-GDM estime la fiabilité de chacune, puis les combine en conséquence. On ferait de même avec deux thermomètres de qualité différente : on accorde plus de poids au plus précis. Sur la figure 8, le réglage trouvé de cette façon tombe au sommet de la courbe, sans utiliser l'image vraie.

FB-GDM est donc une méthode pratique. L'utilisateur fournit seulement la mesure et le modèle de l'appareil : il n'a aucun hyperparamètre à régler et n'a pas besoin de connaître le niveau de bruit. Sur des problèmes de défloutage et d'agrandissement d'images, FB-GDM atteint la qualité de Π GDM réglé à l'aide de l'image vraie [11]. Il reste fiable quand l'appareil ou le niveau de bruit change, et il limite les hallucinations lorsque l'image diffère des images d'apprentissage. Ces propriétés le rendent bien adapté à l'imagerie médicale, où l'image vraie n'est jamais disponible et où les conditions d'examen varient.

DPS, Π GDM et FB-GDM en une phrase

Avec DPS et Π GDM, le dosage entre ce que le modèle a appris et ce que l'appareil a mesuré se règle à la main, souvent à l'aide d'images de référence. Avec FB-GDM, ce dosage est estimé à chaque pas à partir de la mesure seule : aucun réglage n'est nécessaire.

Le prior et l'acquisition sont séparés. Le même modèle de diffusion peut donc servir avec plusieurs appareils ou plusieurs niveaux de bruit. Deux conditions restent nécessaires : les images doivent ressembler au jeu d'entraînement ; l'opérateur A doit aussi représenter correctement l'appareil. Cette approche a été étudiée en IRM et en tomographie [8].

3.4 - Exemples en imagerie médicale

IRM accélérée. L'IRM mesure des coefficients de Fourier dans l'espace k . Mesurer moins de coefficients réduit la durée de l'examen. Mais l'image obtenue directement contient des défauts. Le guidage conserve les coefficients mesurés. Le prior complète les zones manquantes.

Tomographie à faible dose ou à peu de projections. Une dose plus faible ou un plus petit nombre d'angles réduit l'exposition. Mais l'image contient plus de bruit et de stries. L'opérateur A

représente les projections. Le prior complète l'information manquante. Il peut cependant ajouter ou effacer une structure si la mesure est trop pauvre.

Défloutage et dispositifs à coût réduit. Une optique imparfaite ou un mouvement crée du flou. Un noyau peut modéliser ce flou. La reconstruction utilise ce noyau et le prior. Si le noyau est faux, le résultat peut sembler net sans être fidèle.

3.5 - Différence : qualité visuelle et fidélité diagnostique

Un modèle génératif peut produire des images particulièrement réalistes, mais cette capacité présente également des risques. Lorsque les mesures disponibles sont insuffisantes, le modèle peut générer des structures qui ne correspondent pas à la réalité anatomique. Ce phénomène est particulièrement préoccupant en imagerie médicale, car il peut conduire à l'apparition de structures inexistantes ou à la modification de détails cliniquement importants.

La validation doit donc inclure :

- La fidélité dans le domaine des mesures, par exemple l'erreur sur les coefficients IRM acquis ;
- Des métriques d'image telles que PSNR ou SSIM lorsqu'une référence est disponible ;
- Des métriques adaptées à la tâche clinique, comme la détectabilité d'une lésion ;
- Une lecture experte en aveugle ;
- Des tests hors distribution : autre appareil, autre centre, autre protocole ou pathologie rare.

Le PSNR mesure une erreur entre les pixels, il ne mesure pas directement la valeur diagnostique. Une belle image peut encore contenir une erreur locale importante.

Exemple de problème pour la qualité d'images

Prenons l'exemple d'images cérébrales. D'un patient à l'autre, la forme précise des sillons varie beaucoup, mais leur organisation générale reste la même : certaines structures sont toujours situées au même endroit. Un modèle génératif peut très bien apprendre les textures locales du cerveau sans apprendre correctement ces relations spatiales globales. Il peut alors produire une image qui semble réaliste au premier regard, mais qui décrit une anatomie qui n'existe pas. Ce risque rejoint celui de l'hallucination déjà évoqué pour la reconstruction : il concerne aussi la génération pure de nouvelles images.

3.6 - Représentation de l'incertitude

Le processus inverse est aléatoire. On peut donc produire plusieurs reconstructions pour la même mesure. Si elles se ressemblent dans une zone, le résultat est stable. Si elles diffèrent, cette zone reste incertaine.

À partir de K reconstructions $x^{(1)}, \dots, x^{(K)}$, on peut calculer une moyenne et une carte de variance empirique :

$$\bar{x} = \frac{1}{K} \sum_{k=1}^K x^{(k)}, \quad v = \frac{1}{K-1} \sum_{k=1}^K (x^{(k)} - \bar{x})^2. \quad (13)$$

La variance montre les zones qui changent entre les reconstructions. Elle ne donne pas toujours une probabilité exacte d'erreur. Elle aide cependant à repérer les zones sensibles.

4 - Mise en œuvre : de la démonstration au système fiable

4.1 - Architecture et ressources de calcul

Le U-Net reçoit l'image bruitée et l'étape t . Il traite l'image à plusieurs résolutions, ce qui lui permet d'observer la forme générale et les détails. Des blocs d'attention peuvent aussi intégrer une condition.

Un réseau 3D utilise plusieurs coupes en même temps, il permet de mieux décrire le volume, mais nécessite plus de mémoire. Un réseau 2D, lui, est plus léger. Il existe aussi des approches intermédiaires qui utilisent un nombre faible de coupes ou des petits volumes.

L'entraînement des modèles de diffusion nécessite des ressources de calcul importantes et la génération est lente, car elle est effectuée par étapes de débruitage. Certaines méthodes permettent toutefois de réduire ce coût. DDIM diminue le nombre d'étapes nécessaires à la génération, tandis que la diffusion latente effectue le processus dans un espace de représentation compressé plutôt que directement dans l'espace des pixels. Pour donner un ordre de grandeur, générer un patch médical de 64×64 pixels avec ce type de modèle est en pratique impossible sur un simple processeur (CPU) : le calcul prendrait plusieurs mois. La même opération ne prend que quelques secondes sur une carte graphique (GPU) dédiée au calcul. Cet écart explique pourquoi l'accès à une infrastructure de calcul adaptée est une contrainte réelle pour déployer ces méthodes en milieu hospitalier.

4.2 - Protocole expérimental reproductible

Une étude devrait préciser au minimum :

1. L'origine des données, les critères d'inclusion et la répartition par patient ;
2. Les prétraitements, normalisations et résolutions ;
3. L'architecture, le nombre de paramètres et le planning de bruitage ;
4. Le nombre d'étapes d'entraînement, l'optimiseur et les ressources de calcul ;
5. Les méthodes de comparaison et le protocole statistique ;
6. Le temps de calcul et la consommation mémoire à l'inférence ;
7. Les échecs observés, et pas seulement les meilleures images.

Pour une reconstruction, il faut aussi décrire l'opérateur A et le bruit. Leurs paramètres doivent être connus ou estimés. Une méthode peut réussir sur des simulations et échouer sur un appareil réel, c'est pourquoi la simulation doit représenter le système avec assez de précision.

4.3 - Points de vigilance

Risque	Cause possible	Correction possible
Mémorisation	Jeu réduit, surapprentissage	Plus proches voisins, audit de confidentialité, validation externe
Biais de population	Données d'un seul centre ou sous-groupe	Analyse stratifiée, données multicentriques, documentation
Hallucination	Mesure peu informative, guidage trop faible	Contrôle dans le domaine des mesures, cartes d'incertitude

Risque	Cause possible	Correction possible
Artefacts ou instabilité	Planning, solveur ou guidage mal réglés	Études d'ablation, analyse des échecs, tests de sensibilité
Erreur de modèle physique	Opérateur ou bruit mal estimé	Calibration, données réelles, estimation conjointe si nécessaire
Coût de calcul	Grand réseau, nombreuses étapes	Solveur accéléré, compression validée, mesure du temps réel

Table 2 : Principaux risques techniques et pistes de maîtrise

4.4 - Proposition d'activité pédagogique

Ce thème peut faire l'objet d'un travail pratique. Il suffit d'utiliser de petites images, un modèle déjà entraîné et un opérateur simple de flou ou de masquage.

1. Visualiser x_t pour plusieurs valeurs de t et calculer le RSB théorique.
2. Comparer la prédiction du bruit au bruit réellement ajouté.
3. Générer plusieurs images à partir de bruits initiaux différents.
4. Dégrader une image connue selon $y = Ax + b$.
5. Comparer une inversion naïve, une régularisation classique et un guidage par diffusion.
6. Faire varier la force de guidage et relever simultanément l'erreur de mesure et l'erreur à l'image de référence.
7. Produire plusieurs reconstructions et observer la carte de variance.

Ce TP relie le traitement du signal, les probabilités et les réseaux de neurones. Il montre surtout qu'une image réaliste n'est pas toujours fidèle à la mesure.

À titre d'illustration, un tel apprentissage sur de petites images de chiffres manuscrits (résolution 28×28) permet déjà d'obtenir des images reconnaissables après un léger entraînement sur une carte graphique standard, avec toutefois quelques artefacts sur certains échantillons.

5 - Conclusion

Les modèles de diffusion répondent à deux besoins en imagerie médicale. Ils peuvent créer des données pour compléter un jeu d'apprentissage. Ils peuvent aussi servir de prior pour reconstruire une image.

Le principe tient en trois idées. Un processus connu ajoute du bruit aux images. Un réseau apprend ensuite à retirer ce bruit. Enfin, une condition ou une mesure peut guider le processus inverse.

La diffusion n'est cependant pas la seule approche pour transformer un bruit simple en une image complexe. Les normalizing flows poursuivent le même objectif par une transformation directement inversible, ce qui permet de calculer une probabilité exacte et de générer une image en une seule étape. Cette méthode reste aujourd'hui moins utilisée que la diffusion sur des images complexes, mais elle a inspiré une évolution récente, le flow matching, qui combine la rapidité des flows et la qualité de la diffusion. Cette proximité entre les deux familles de modèles montre qu'elles répondent, par des chemins différents, à une même question : comment transformer une distribution simple en une distribution aussi complexe qu'un jeu d'images médicales.

La validation reste essentielle, quelle que soit la méthode choisie. Une image synthétique doit améliorer une tâche testée sur des données réelles. Une reconstruction doit respecter les mesures du patient. Elle doit aussi conserver l'information utile au diagnostic.

Les progrès à venir ne concernent donc pas seulement la qualité des images. Ils concernent aussi la vitesse, l'incertitude, la confidentialité et la robustesse face à un nouvel appareil ou un nouveau protocole. Sur ce point, des approches comme le flow matching pourraient réduire le temps de calcul, un enjeu particulièrement important pour un déploiement en milieu hospitalier.

Bibliographie

- [1] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan et S. Ganguli, « Deep Unsupervised Learning using Nonequilibrium Thermodynamics », *Proceedings of ICML*, 2015.
- [2] J. Ho, A. Jain et P. Abbeel, « Denoising Diffusion Probabilistic Models », *Advances in Neural Information Processing Systems*, 2020.
- [3] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon et B. Poole, « Score-Based Generative Modeling through Stochastic Differential Equations », *International Conference on Learning Representations*, 2021.
- [4] A. Q. Nichol et P. Dhariwal, « Improved Denoising Diffusion Probabilistic Models », *Proceedings of ICML*, 2021.
- [5] J. Song, C. Meng et S. Ermon, « Denoising Diffusion Implicit Models », *International Conference on Learning Representations*, 2021.
- [6] J. Ho et T. Salimans, « Classifier-Free Diffusion Guidance », *NeurIPS Workshop on Deep Generative Models and Downstream Applications*, 2021.
- [7] O. Ronneberger, P. Fischer et T. Brox, « U-Net: Convolutional Networks for Biomedical Image Segmentation », *Medical Image Computing and Computer-Assisted Intervention*, 2015.
- [8] Y. Song, L. Shen, L. Xing et S. Ermon, « Solving Inverse Problems in Medical Imaging with Score-Based Generative Models », *International Conference on Learning Representations*, 2022.
- [9] H. Chung, J. Kim, M. T. McCann, M. L. Klasky et J. C. Ye, « Diffusion Posterior Sampling for General Noisy Inverse Problems », *International Conference on Learning Representations*, 2023.
- [10] J. Song, A. Vahdat, M. Mardani et J. Kautz, « Pseudoinverse-Guided Diffusion Models for Inverse Problems », *International Conference on Learning Representations*, 2023.
- [11] G. Ségué et T. Rodet, « FB-GDM: Fully-Bayesian Guided Diffusion Models for High-Dimensional Linear Inverse Problems via Unsupervised Variational Inference », *arXiv:2609.29216*, 2026.
- [12] A. Kazerouni et al., « Diffusion Models in Medical Imaging: A Comprehensive Survey », *Medical Image Analysis*, vol. 88, 2023.
- [13] D. J. Rezende et S. Mohamed, « Variational Inference with Normalizing Flows », *Proceedings of ICML*, 2015.
- [14] L. Dinh, J. Sohl-Dickstein et S. Bengio, « Density Estimation using Real NVP », *International Conference on Learning Representations*, 2017.
- [15] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed et B. Lakshminarayanan, « Normalizing Flows for Probabilistic Modeling and Inference », *Journal of Machine Learning Research*, vol. 22, 2021.
- [16] R. T. Q. Chen, Y. Rubanova, J. Bettencourt et D. Duvenaud, « Neural Ordinary Differential Equations », *Advances in Neural Information Processing Systems*, 2018.