

# 1 摘要

本計畫擬提出一套基於條件式擴散模型之三維肺部 CT 病灶合成方法，目標為提升生成影像之紋理真實性。模型以真實背景影像與病灶遮罩為條件，限制生成於前景區域內，以維持結構一致性。方法基於 Denoising Diffusion Probabilistic Models (DDPM) 改良，擴散模型之由高雜訊至低雜訊之階段性演化特性本身即隱含由低頻結構逐步細化至高頻紋理之生成過程，本研究基於此特性，設計具時間條件化 (time-conditioned) 之多階段損失加權策略，以提升模型於不同生成階段對應之學習目標對齊能力。

既有研究指出，僅以均方誤差 (Mean Squared Error, MSE) 訓練易導致過度平滑，難以重建具感知與語意意義之高頻細節；醫學影像判讀又高度依賴中高頻紋理與邊界特徵，因此單一像素層級損失不足以保留診斷關鍵資訊。雖有研究主張依時間步混合多重損失可提升品質，本計畫前期實驗顯示，即使結合信噪比加權與差分高斯 (Difference of Gaussian, DoG) 高頻濾波，仍未顯著改善紋理品質，顯示通用高頻強化不足以捕捉醫學語意紋理。

因此，本計畫將嘗試引入具醫學紋理意義之特徵導向損失，包括多頻多方向 Gabor 紋理損失、可微分灰階共生矩陣 (Gray-Level Co-occurrence Matrix, GLCM) 分佈對齊損失，以及基於離散小波轉換 (Discrete Wavelet Transform, DWT) 之高頻子頻帶能量加權損失，並於擴散後期啟用高頻約束以強化細節重建能力，提升生成影像之紋理一致性與臨床可判讀性。

## 2 研究動機與研究問題

### 2.1 研究動機

#### 2.1.1 醫學影像標註資料稀缺與合成需求

在醫學影像分析任務中，具高品質標註之訓練資料長期處於稀缺狀態。相較於自然影像，醫學影像的資料蒐集與標註高度仰賴專業醫師判讀，不僅成本高昂，亦受限於臨床工作負荷與倫理規範，使得大規模、細緻標註資料難以取得。此外，許多具臨床價值之病灶類型在實際病例中本就屬於低發生率事件，進一步造成資料分佈不平衡的問題，限制深度學習模型於實務應用中的泛化能力。

在此背景下，利用生成式模型合成具臨床合理性之醫學影像，已成為緩解標註資料稀缺、提升下游模型訓練穩定性與泛化能力之重要研究方向。相較於傳統生成式對抗網路，擴散模型 (Diffusion Models) 將影像生成過程形式化為具明確機率語義的逐步去噪過程，其訓練目標可由變分下界或等價之分數匹配觀點推導而得，使生成機制與最佳化目標在機率層級上具有清楚的數學詮釋。此一特性使擴散模型在訓練穩定性、樣本多樣性，以及生成過程可分析性方面，相較對抗式學習架構更具結構性優勢。

基於上述特性，擴散模型近年逐漸被引入醫學影像合成任務，並於多項研究中展現出相較傳統方法更穩定且一致的生成表現。然而，現行擴散模型多源自自然影像生成脈絡，其訓練目標與設計假設主要針對一般視覺品質與像素分佈對齊進行優化，未必能充分反映醫學影像在臨床判讀中所重視之組織特異性紋理、局部結構差異與病灶判別特徵。因此，如何調整擴散模型之學習目標與生成機制，使其更有效對齊醫學影像之臨床需求，仍有待進一步系統性探討。

#### 2.1.2 像素空間擴散模型於醫學影像合成之侷限

近年來，擴散模型 (Diffusion Models) 已廣泛應用於高品質影像生成，其中以 DDPM 為代表之像素空間模型多以均方誤 (MSE) 作為主要訓練目標 (Ho, Jain, and Abbeel [1])。然而，MSE 僅衡量像素層級之亮度差異，缺乏對結構、語意與感知層級差異的刻畫能力，因此在主觀視覺品質與人類感知一致性上存在限制 (An et al. [2])。因臨床判讀仰賴具組織特異性之細微紋理與結構模式而非單純像素相似度，此問題在醫學影像生成中更為明顯。

此外，傳統擴散模型於所有時間步長均採用同形式與權重之 MSE，隱含不同擴散階段具有相同資訊尺度的假設。然而生成過程實際具階段性：高雜訊階段偏向學習低頻與全域結構，低雜訊階段則負責高頻細節與微結構重建。若忽略此時間尺度差異，後期紋理重建可能監督不足，影響最終細節品質 (Hang et al. [3])。

基於此，本計畫前期實驗嘗試結合 MSE 與 DoG 損失，以高頻導向誤差補強邊緣與細節。然而實驗結果顯示整體提升有限，未較原始 DDPM 架構產生穩定且顯著改善 (4.2.2)。

其原因在於，DoG 僅刻畫局部一階導數結構，而醫學影像判讀涉及二階以上紋理統計與結構關係，例如空間共現、方向一致性與尺度耦合等高階特徵。由於此類資訊未被明確建模，DoG 所提供之監督訊號在表示層級上與臨床判別依據存在落差，即使提高其損失權重亦難以彌補。

因此，儘管時間步加權或多損失混合在理論上具合理性 (3.3)，若缺乏具病灶判別力之紋理表示與對應監督訊號，僅調整權重分配仍難以實質提升模型對病灶紋理真實性的建模能力，從而限制生成品質的進一步改善。

## 2.2 研究目的

本計畫之核心研究目標在於：依據擴散模型於不同噪聲時間步 (timestep) 下所對應之生成行為差異，設計具時間條件化 (time-conditioned) 之學習目標與特徵導向損失函數，以提升醫學影像生成之紋理真實性與臨床可判讀性。

具體而言，在高噪聲之 early steps 階段，生成過程主要對應於低頻與整體結構資訊之建立，因此以結構導向之學習目標為主，以確保病灶整體形態與空間配置之穩定形成；而於低噪聲之 late steps 階段，模型逐步補足高頻細節，故引入專為醫學影像紋理特性設計之特徵型損失函數，例如基於方向與頻率分析之 Gabor 濾波特徵，或基於區域灰階統計之灰階共生矩陣 (GLCM) 指標，以彌補僅依賴梯度資訊之 DoG 特徵在紋理表徵上的不足。

此外，本計畫亦考慮採用編碼-解碼 (encoder-decoder) 架構將醫學影像由像素空間映射至潛在表示空間，並於該潛在空間中進行擴散建模。相較於直接於像素空間進行生成，此一表示轉換可降低逐像素誤差對生成過程的主導性，使模型在建模高頻細節與紋理結構時具備更高的彈性。

透過此一依時間步條件化之特徵導向損失設計，本計畫期望在不犧牲結構穩定性的前提下，使模型更精準地學習醫學影像中具診斷價值之紋理模式，進而提升生成影像於臨床應用情境中的可信度與實用性，並為像素空間擴散模型於醫學影像生成任務中的損失函數設計提供具體且可重現之改良方向。

## 3 文獻回顧與探討

### 3.1 利用生成式模型合成具紋理特徵之醫學影像

相較於自然影像，醫學影像的生成任務具有更為嚴格的結構與紋理要求，其生成結果須具備足夠的臨床可判讀性與診斷價值。因此，相關研究多著重於如何設計能反映醫學影像特性的專用損失函數，以提升生成影像在紋理真實性與結構一致性上的表現。本實驗室先前研究 (Huang [4]、Chen et al. [5]) 已成功引入基於 Gabor filter 之紋理特徵損失函數，透過多尺度與多方向濾波回應來捕捉醫學影像中的空間頻率與方向性紋理資訊，並於肺結節影像生成任務中有效改善生成影像的細節紋理真實性。

除 Gabor-based 方法外，Yuan et al. [6] 進一步提出一種基於灰階共生矩陣 (GLCM) 的紋理損失函數，將傳統用於放射組學分析的二階統計紋理特徵整合至生成模型的訓練流程中。該方法透過計算生成影像與真實影像在 GLCM 特徵空間中的差異，直接約束影像在特定區域內的紋理統計分布，從而引導模型學習更符合臨床影像特性的紋理表現。實驗結果顯示，所提出之 GLCM 紋理損失能在不犧牲整體影像品質的情況下，顯著提升生成影像於紋理一

致性與放射組學特徵可重現性上的表現，進一步驗證了針對醫學影像特性所設計之紋理損失函數，在醫學影像生成任務中的關鍵角色。

近年來，潛在擴散模型（Latent Diffusion Models, LDM）亦被用以強化高解析度影像之高頻細節重建，例如 Sigillo, He, and Comminiello [7] 提出之 Latent Wavelet Diffusion，透過小波轉換將影像分解為多頻段表示，並於潛在空間中進行擴散建模。在醫學影像領域中，LDM 多半被用以降低三維高解析資料之計算成本，如 Wang et al. [8]，但亦有研究進一步利用潛在空間之特徵建模能力以改善紋理生成，例如 Yin et al. [9] 於共享潛在空間中引入由 CBCT 萃取之高頻資訊作為條件以引導去噪過程、Jiang et al. [10] 透過小波轉換顯式分離低頻與高頻成分。總而言之，相關研究顯示，結合潛在擴散模型與頻率或高頻導向設計，有助於同時兼顧計算效率與醫學影像生成之紋理真實性。

### 3.2 MSE 作為擴散模型訓練目標之侷限性

擴散模型將影像生成過程形式化為具明確機率語義的逐步去噪過程。其訓練目標可由變分下界或等價之分數匹配觀點推導而得，使生成機制與最佳化目標在機率層級上具有清楚的數學詮釋。基於此一嚴謹的機率建模框架，擴散模型已成為當前主流之影像生成方法。

然而，儘管其理論基礎完整，傳統擴散模型於實作上多採用均方誤（MSE）作為訓練目標。多項研究指出，此一像素層級的重建誤差在感知品質對齊上存在根本限制。An et al. [2] 指出，MSE 僅衡量逐點差異，難以反映影像於結構與語意層級的差異，因此無法充分對齊人類主觀視覺品質。Mathieu, Couprie, and LeCun [11] 亦觀察到，在生成任務中若僅以 MSE 作為最佳化目標，模型往往產生過度平滑與細節模糊之結果。

此一現象可由 Blau and Michaeli [12] 所提出之 perception-distortion tradeoff 理論加以解釋。該理論指出，在任何以失真度量（如 MSE）為最佳化目標的重建系統中，感知品質與失真誤差之間存在不可避免的權衡關係。當真實資料分佈具有多模態特性時，MSE 對應之 Bayes optimal 解將趨近於後驗平均（posterior mean）。換言之，模型實際學習的是條件期望，而非條件分佈本身。此種後驗平均雖能有效降低像素誤差，卻會對多種可能解進行平均化，導致生成結果偏離自然影像流形，並產生邊緣鈍化與高頻紋理流失等現象。

除機率層級之解釋外，頻率域分析亦提供另一種理解途徑。Rahaman et al. [13] 指出，神經網路在梯度下降訓練過程中存在顯著的頻譜偏好（spectral bias），即模型傾向優先擬合低頻成分，而高頻成分則學習速度較慢。當此一頻譜偏好與 MSE 所隱含的平均化特性同時存在時，模型在優化過程中將更進一步偏向低頻結構的重建，而對感知品質至關重要的高頻細節則被系統性弱化。兩者的疊加效應，使生成影像在主觀視覺上呈現模糊且缺乏真實感。

上述限制在醫學影像生成任務中尤為嚴重。臨床判讀高度依賴細微紋理與邊界形狀，而此類資訊通常對應於中高頻成分。若僅以 MSE 作為統一訓練目標，模型在其內在頻譜偏好與 perception-distortion tradeoff 的共同作用下，將難以充分重建臨床關鍵細節。因此，有必要引入能直接約束頻域行為或感知品質之輔助目標，以提升醫學影像生成在結構保真度與紋理真實性上的表現。

### 3.3 多權重混合之擴散模型

Hang et al. [3] 從時間尺度的角度指出，傳統擴散模型於訓練時對所有時間步長  $t$  一律採用相同形式與權重之 MSE 作為學習目標，隱含假設不同擴散步驟所對應之重建任務具有相同的資訊尺度。然而實際上，高雜訊時間步主要承載影像之低頻結構與全域配置，而低雜訊時間步則負責高頻細節與紋理資訊之重建；若忽略此一差異，將導致模型於特定時間區段之學習不足。為此，該研究提出以時間步為基礎之損失重加權策略（step-wise loss reweighting），以更精確地對齊不同噪聲尺度下的學習需求。

類似的階段性生成概念亦可見於多尺度架構設計中。Xiao et al. [14] 提出金字塔式多尺度生成架構（pyramid hierarchical structure），由低解析度層級先行建模影像之整體結構，再逐步於高解析度層級細化局部細節與紋理，使不同尺度之生成任務得以分階段進行。此一金字塔層級生成策略，在概念上可視為擴散模型時間維度結構化之空間對應：低解析度層級對應

高雜訊階段之全域結構建模，而高解析度層級則對應低雜訊階段之高頻細節重建，顯示階段性生成策略有助於提升整體生成品質。

在上述脈絡下，Ma, Xu, and Zhang [15] 進一步將時間尺度差異具體化為多損失函數之階段性混合設計。該研究明確指出，像素空間擴散模型於不同噪聲時間步所對應之生成任務，具有顯著的感知層級差異：高雜訊階段側重整體結構與語意配置，而低雜訊階段則承載高頻紋理與視覺細節。因此，該方法僅於低雜訊時間區段啟用 LPIPS 與基於自監督特徵之 P-DINO 感知損失，以避免過早施加感知偏置而損及樣本多樣性。此種以時間步為條件之多損失混合策略，進一步驗證擴散模型之生成歷程具有內在的結構化階段性，需依不同噪聲層級配置相符之學習目標，方能同時兼顧生成品質與感知一致性。

## 4 研究方法及步驟

### 4.1 資料來源與前置處理

本計畫所使用之資料集為 Armato 等人所建立之 LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) 資料庫 (Armato et al. [16])，並由 The Cancer Imaging Archive 提供公開存取 (National Cancer Institute [17])。LIDC-IDRI 資料集包含超過 1000 位患者的肺部 CT 掃描影像，並由多位放射科醫師進行結節標註，提供豐富的肺部結節影像資料。本計畫將從 LIDC-IDRI 資料集中篩選出包含肺結核病灶的 CT 影像，並進行前置處理以符合研究需求。

#### 4.1.1 資料分類

LIDC-IDRI 資料集中的 CT 影像格式為 DICOM (Digital Imaging and Communications in Medicine)，每位患者的 CT 掃描除了多張 dicom 切片外，還包含對應的結節標註檔案 (XML 格式)。標註檔內容為一到四位放射科醫師對該患者 CT 影像中結節的標註結果，分別為  $\geq 3$  mm、 $< 3$  mm 以及非結節三類。

放射科醫師會對大於等於 3 mm 的結節進行進一步的診斷，並根據結節的惡性程度給予 1 至 5 分的評分 (1 分為高度良性 (Benign)，5 分為高度惡性 (Malignancy))。由於多位放射科醫師會對同一患者的 CT 影像進行標註，因此同一結節可能會有多個評分結果。本計畫將採用所有放射科醫師對同一結節評分的平均值作為該結節的最終評分，並且選取評分為 1 或 2 分的結節作為良性肺結核病灶，4 或 5 分的結節作為惡性肺結核病灶，總共選取 883 筆良性以及 492 筆惡性肺結核病灶進行後續實驗。

#### 4.1.2 資料處理

本計畫沿用實驗室學長於早期研究 (Chen [18]) 中所建立之訓練資料集。該研究已完成肺部區域分割以及結節 ROI (Region of Interest) 的裁切，並將結果儲存為尺寸為  $64 \times 64 \times 64$  的 .mat 格式檔案，其中 CT 影像之 Lung Window 設定為 Window Width = 1500、Window Level = -400。

為提升資料讀取效率並方便後續深度學習模型訓練，本計畫將原始 .mat 檔案轉換為 .npy 格式作為主要資料儲存形式。在資料轉換過程中，本計畫自原始 .mat 檔案中擷取 Seg\_VOI 與 TDM\_VOI 兩個欄位，分別代表肺部結節之二值分割遮罩 (segmentation mask) 以及結節 CT 影像區域 (Tumor Density Mask)。上述兩項資料分別作為模型訓練時的標註資訊與輸入影像，以建立結節辨識與分析之深度學習資料集。

## 4.2 前期研究

### 4.2.1 前期研究方法

本計畫之前期研究旨在驗證「多階段損失加權」是否能提升肺結核病灶 CT 影像生成之紋理真實性與細節還原能力。基於擴散模型由低頻結構逐步細化至高頻紋理的生成特性，假設：若能依 diffusion timestep 動態調整不同頻率層級之監督強度，模型可更有效地學習病灶形態與內部紋理結構。前期研究採用簡化之條件式擴散模型，依循 DDPM 去噪學習框架，並以 U-Net (Ronneberger, Fischer, and Brox [19]) 為骨幹網路，結合時間步長嵌入與殘差模組建構反向擴散預測函數。參考 Nichol and Dhariwal [20]，模型直接預測乾淨影像  $x_0$ ，並於該空間施加低頻與高頻重建損失。輸入包含前景遮罩  $x_{mask}$ 、原始影像  $x_0$  與混合雜訊之影像  $x_t$ ，輸出為預測之乾淨影像  $\hat{x}_0$ ，架構如圖 1 所示。

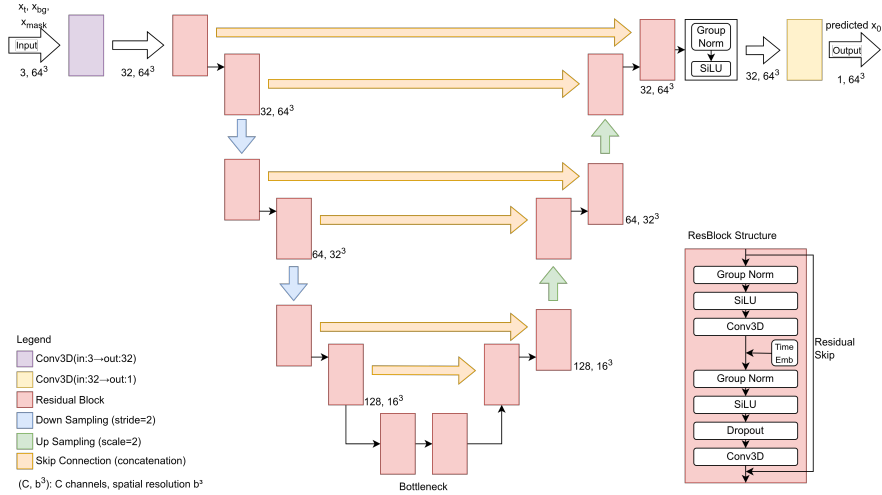


Figure 1: 前期研究模型架構圖

由於 Malignancy 類別的資料量較少，訓練速度較快，因此在前期實驗中主要使用 Malignancy 類別的資料進行模型訓練與評估。由於資料量並不多，因此本計畫以 1:9 的比例切分為測試集（Test set）以及訓練集（Training set）。其中訓練集包含 443 筆惡性肺結核病灶資料，而測試集則包含 49 筆惡性肺結核病灶資料。訓練時模型僅會透過訓練集資料進行學習，測試集資料則完全不參與訓練過程，僅用於最終評估模型性能，以確保評估結果的客觀性與泛化能力。

在此基礎上，引入低頻重建損失  $\mathcal{L}_{low}$  (MSE) 與高頻重建損失  $\mathcal{L}_{high}$  (DoG)，並設計四種具時間步長感知（step-aware）特性之多階段損失加權策略：

#### 1. SNR-based weighting

利用擴散過程中信噪比（Signal-to-Noise Ratio, SNR）隨時間步長  $t$  增加而提升的特性，動態調整高頻損失權重，使其在低 SNR 時受到抑制、高 SNR 時逐漸增強。定義

$$w_{high}(t) = \frac{SNR(t)}{SNR(t) + 1}, \quad SNR(t) = \frac{\bar{\alpha}_t}{1 - \bar{\alpha}_t + \epsilon}, \quad (1)$$

其中  $\bar{\alpha}_t$  為累積噪聲衰減係數， $\epsilon$  為數值穩定項。整體損失函數定義為

$$\mathcal{L}(t) = \mathcal{L}_{low} + \lambda w_{high}(t) \mathcal{L}_{high}, \quad (2)$$

其中  $\lambda$  為控制高頻重建損失之相對權重，前期實驗中固定設為 0.3。

## 2. FiLM-based weighting

參考 Perez et al. [21]，將時間步長  $t$  視為條件訊號，透過一條條件網路  $g_\phi$  學習時間相關之調制參數，並以 FiLM 型線性調制結合 Sigmoid 函數，生成高頻損失之時間相關權重：

$$(\gamma(t), \beta(t)) = g_\phi(t), \quad w_{\text{high}}(t) = \sigma(\gamma(t)t + \beta(t)), \quad w_{\text{high}}(t) \in [0, 1], \quad (3)$$

總損失函數定義為

$$\mathcal{L}(t) = (1 - w_{\text{high}}(t))\mathcal{L}_{\text{low}} + w_{\text{high}}(t)\mathcal{L}_{\text{high}}. \quad (4)$$

該方法允許模型以資料驅動方式，自動學習在不同擴散階段中高頻監督的重要性。

## 3. Uncertainty-based weighting

參考 Kendall, Gal, and Cipolla [22] 提出之同質不確定性 (homoscedastic uncertainty) 建模方法，將低頻與高頻重建視為兩個子任務，分別引入可學習之不確定性參數  $\sigma_{\text{low}}$  與  $\sigma_{\text{high}}$ 。整體損失函數定義為

$$\mathcal{L} = \frac{1}{2\sigma_{\text{low}}^2} \mathcal{L}_{\text{low}} + \frac{1}{2} \log \sigma_{\text{low}}^2 + \frac{1}{2\sigma_{\text{high}}^2} \mathcal{L}_{\text{high}} + \frac{1}{2} \log \sigma_{\text{high}}^2. \quad (5)$$

藉由學習不確定性參數，模型可自動降低高不確定性 (較難學習) 任務之影響，並提升相對穩定任務的權重，以達到多任務訓練平衡。

## 4. Step-aware control

此方法並非根據時間步長調整不同損失函數比例，而是直接依據時間步長調整監督區域的範圍。於生成初期 ( $t < T/2$ ,  $t$  代表當前步數,  $T$  為總步數)，僅在經侵蝕處理之前景區域內計算重建誤差，使模型專注於整體結構學習；於生成後期 ( $t \geq T/2$ )，恢復對完整前景區域的監督，以強化邊界細節與局部一致性。雖然此策略僅使用 MSE 作為損失函數，但透過時間感知的監督範圍調整，概念上亦為一種階段性學習機制，使模型在不同生成階段專注於不同層級的特徵學習，進而提升整體生成品質。

上述方法本質上皆屬於「多階段損失加權」之不同實作形式，其核心假設為：DoG 能有效捕捉醫學影像之紋理特徵，透過時間或階段相關設計，可使模型逐步由低頻形態學習過渡至高頻紋理建模，進而提升生成影像整體真實性。

除了以上的實驗外，本計畫在前期實驗中亦訓練了一個以生成對抗神經網路 (Generative Adversarial Network, GAN) 為基礎的模型，作為與擴散模型進行比較之基準方法。該 GAN 模型同樣以背景影像與前景遮罩作為條件，學習生成對應之肺結核病灶 CT 影像。透過對比分析，我們期望能更全面地評估擴散模型在醫學影像生成任務中的優勢與不足。

### 4.2.2 前期研究結果

圖 2 分別展示了上述實驗在 Test Set 之生成結果比較，其中 (a) 為真實影像，(b) 為前景遮罩，(c) - (h) 分別對應 DDPM、SNR-based weighting、FiLM-based weighting、Uncertainty-based weighting、Step-aware control 以及 GAN 方法。

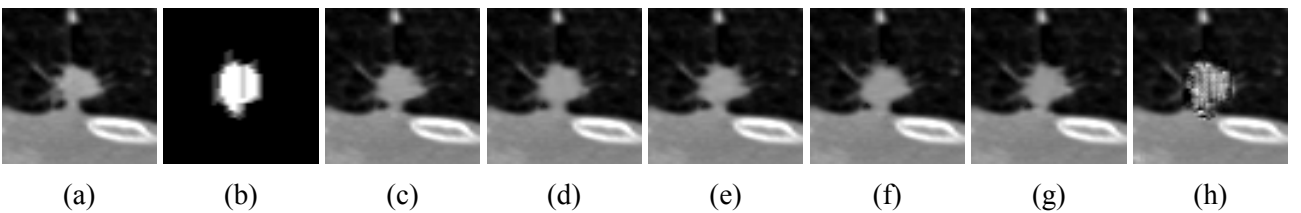


Figure 2: 不同生成方法於相同背景與前景遮罩條件下之病灶生成結果比較 (Case 2)，(a) 真實影像，(b) 前景遮罩，(c) DDPM，(d) SNR-based weighting，(e) FiLM-based weighting，(f) Uncertainty-based weighting，(g) Step-aware control，(h) GAN



視覺上觀察，擴散模型（DDPM）在病灶區域的紋理細節與整體結構上均較 GAN 模型（GAN）表現出更高的真實性與合理性。其他方法皆採用擴散模型之架構進行改良，生成結果極為相似，肉眼幾乎無法區分。因此，本計畫同時也使用紋理分析方法對生成結果進行定量評估，以客觀比較不同實驗之紋理真實性，具體方法將於下一節4.3.3說明。

圖3放大檢視生成結果，比較 baseline DDPM 與真實影像之差異。觀察可見，DDPM 生成結果在病灶區域之細部紋理資訊明顯少於原始影像，顯示其在高頻細節上仍有不足之處。

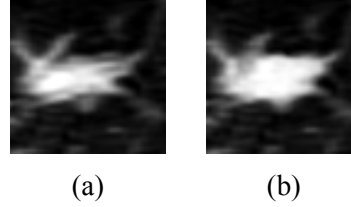


Figure 3: 真實影像與 DDPM 生成結果之細節比較，(a) 真實影像，(b) DDPM 生成結果

表 1 彙整了前期實驗中各組生成資料與 Test Set 之比較結果，生成結果係以 Test Set 之背景作為引導條件，並在前景區域內生成病灶 CT 影像，比較內容皆僅包含生成的前景。為更直觀展現生成結果，MSE 以  $10^{-2}$  為單位呈現，且表中所有指標皆以平均值與括號內之標準差表示。綠色字體表示在該指標上表現最佳之方法，紅色字體則表示表現最差之方法，WD 代表 Wasserstein Distance。

	GLCM WD(std)	GLRLM WD(std)	Gabor WD(std)	FID	MSE ( $\times 10^{-2}$ )
DDPM	0.016(0.012)	6.54(14.49)	0.00058(0.00096)	22.40	0.028(0.047)
SNR	0.017(0.012)	6.83(14.88)	0.00063(0.00085)	27.67	0.023(0.040)
FiLM	0.017(0.013)	6.03(13.14)	0.00061(0.0010)	22.20	0.028(0.048)
Uncertainty	0.020(0.018)	6.93(15.27)	0.00077(0.0013)	25.56	0.035(0.060)
Step-aware	0.017(0.012)	6.39(13.84)	0.00061(0.00084)	25.09	0.026(0.042)
GAN	0.14(0.17)	54.48(122.55)	0.0035(0.0029)	100.88	792.0(2355.0)

Table 1: 前期實驗結果與 Test Set 之比較

由表 1 可觀察到，所有基於 DDPM 架構之模型在各評估指標上皆顯著優於生成對抗網路（GAN）。此結果顯示，GAN 在肺結核病灶 CT 影像生成任務中難以同時維持病灶區域之紋理一致性與整體結構穩定性；相較之下，DDPM 架構能夠穩定地生成紋理特徵與全域分佈上均更貼近真實資料的影像，驗證了擴散模型在高保真醫學影像生成任務中之結構優勢與可行性，亦構成本計畫後續方法設計的基礎。

進一步分析基於 DDPM 架構之多種多階段損失加權策略可發現，儘管部分方法在特定指標上（例如 FiLM-based weighting 於 GLCM WD 與 FID）相較原始 DDPM 略有改善，整體而言仍未呈現出顯著且一致的性能優勢；部分方法（如 Uncertainty-based weighting）甚至在多項評估指標上表現劣於 baseline。此類方法之核心設計僅對於既有損失函數隨時間步長進行權重調整，卻未從根本上提升損失函數對病灶紋理特徵之表徵能力，因此其對生成品質之影響多半侷限於數值層面的重新分配，而非實質性之特徵學習增益。

## 4.3 研究方法

### 4.3.1 模型架構與改良方法

在前期研究中，為提昇肺結核病灶 CT 影像生成之紋理真實性與細節還原能力，已嘗試引入多種「多階段損失加權」之改良方法，包含基於信噪比之損失加權、基於 FiLM 調制之損失加權等設計。然而，上述方法在定量評估與視覺化分析中，並未展現出相對於原始 DDPM

模型之顯著優勢，此結果顯示單純調控生成過程中不同時間尺度之監督訊號，並不足以有效提升肺結核病灶之紋理真實性。

考量在前期研究中使用之兩種損失函數（MSE, DoG）皆非專門針對紋理特徵設計，且未明確引入具判別力之病灶特徵描述。綜合上述結果，本計畫擬提出一套更具針對性且專門設計之改良方法，希冀其能在 early steps 階段保留病灶之整體結構特徵，並在 late steps 階段強化病灶之紋理細節還原能力。

為有效捕捉醫學影像中具診斷價值之紋理特徵，本計畫將引入幾種專門設計之紋理損失函數，並結合多階段損失加權策略，以期在不同生成階段對病灶特徵進行更精確的約束與引導。具體而言，本計畫擬引入以下幾種損失函數：

1. Gabor filter-based texture loss：為量化生成影像與真實影像在空間頻域紋理特徵上的相似性，本計畫參考實驗室先前研究成果 Huang [4]、Chen et al. [5]，設計基於 Gabor filter 之紋理損失函數。Gabor filter 主要由正弦波與高斯函數組成，能有效捕捉影像中的空間頻率與方向性紋理特徵（Manjunath and Ma [23]）。

首先，對生成影像與真實影像分別套用一組二維 Gabor filters，以擷取不同空間頻帶下之紋理回應。考量輸入資料為三維 CT 影像，且直接使用三維 Gabor filter 會導致計算成本與參數複雜度顯著增加，本計畫採用 multi-planar slice-wise 策略：分別沿  $xy$  (axial)、 $xz$  (coronal) 與  $yz$  (sagittal) 平面對三維影像逐張切片施加二維 Gabor 卷積，最終將三平面取得之 loss 結果平均，以獲得整體紋理相似性評估。

對生成影像以及真實影像做完 Gabor 卷積後，計算兩者在各頻率與方向組合下 response 之 MSE，並以病灶區域之二值遮罩作為加權因子，確保損失僅於結節區域內計算，以避免背景區域之雜訊對紋理學習造成干擾。

具體而言，本計畫之 Gabor loss 定義如下，其中， $g_{(p,s)}$  與  $r_{(p,s)}$  分別代表生成與真實樣本之前景區域， $p$  代表像素索引， $s$  代表 Gabor filter 之索引； $f$  代表 Gabor filter， $N_p$  與  $N_s$  分別為 VOI 中每張 slice 具有之 pixel 總數與 slice 總數， $N_f$  代表 Gabor filter 總數； $w$  為可訓練之 filter 權重參數

$$\mathcal{L}_{\text{Gabor}} = \frac{1}{N_f} \sum_{s=1}^{N_s} \sum_{p=1}^{N_p} \left( (f * g_{(p,s)} - f * r_{(p,s)})^2 \times w \right), \quad (6)$$

2. Co-occurrence matrix-based loss：參考 Yuan et al. [6] 所提出之灰階共生矩陣（GLCM）損失函數，透過直接對齊合成影像與真實影像在灰階共生分佈上的差異，以顯式約束影像之空間共現關係，進而提升紋理重建品質。

給定真實影像  $x$ 、模型預測影像  $\hat{x}$  以及前景遮罩  $m$ ，首先以可微分之軟量化（soft binning）方式，將灰階值映射至  $K$  個分箱，其對第  $k$  個分箱之權重定義為

$$w_k(v) = \frac{\exp\left(-\frac{(v-c_k)^2}{\tau^2}\right)}{\sum_{\ell} \exp\left(-\frac{(v-c_{\ell})^2}{\tau^2}\right)}. \quad (7)$$

其中， $c_k$  為第  $k$  個分箱之中心值， $\tau$  為控制分箱寬度的參數。接著，對於每一空間位移方向  $\Delta$ ，僅在前景區域內計算體素對之共生矩陣：

$$C_{ij}^{(\Delta)}(x) = \sum_{\mathbf{p}} m(\mathbf{p}) m(\mathbf{p} + \Delta) w_i(x(\mathbf{p})) w_j(x(\mathbf{p} + \Delta)), \quad (8)$$

並進一步正規化為共生機率分佈  $P^{(\Delta)}(x)$ 。最終，定義共生矩陣損失為多方向共生分佈差異之平均：

$$\mathcal{L}_{\text{CM}} = \frac{1}{|\mathcal{D}|} \sum_{\Delta \in \mathcal{D}} \|P^{(\Delta)}(\hat{x}) - P^{(\Delta)}(x)\|_1. \quad (9)$$



3. Wavelet-based texture loss：參考 Jiang et al. [10]，本計畫於 Latent Diffusion Model (LDM) 架構中引入一種基於小波高頻子頻帶能量加權之紋理損失函數 (wavelet-based texture loss)，以增強模型對細緻紋理與結構邊界的重建能力。

離散小波轉換 (DWT) 可將影像分解為一低頻近似子頻帶 ( $LL$ ) 及三個具方向性的高頻細節子頻帶：水平 ( $LH$ )、垂直 ( $HL$ ) 與對角 ( $HH$ )。由於高頻子頻帶主要對應邊緣與局部紋理，本計畫在紋理相似性評估時僅考慮高頻子頻帶，以避免低頻亮度與整體結構差異主導損失。

對生成影像  $x$  與真實影像  $y$  施加二維 DWT，僅考慮高頻子頻帶  $b \in LH, HL, HH$ ，其小波係數表示為  $W_b(\cdot)$ 。根據真實影像高頻能量構建空間權重圖  $R$ ：

$$R = \frac{\sum_{b \in LH, HL, HH} |W_b(y)|^2}{\max \left( \sum_{b \in LH, HL, HH} |W_b(y)|^2 \right)}, R \in [0, 1] \quad (10)$$

最終 wavelet-based texture loss 定義為

$$\mathcal{L}_{\text{Wavelet}} = \sum_{b \in LH, HL, HH} (\alpha + \beta R) \odot \rho(|W_b(x) - W_b(y)|), \rho(z) = \sqrt{z^2 + \epsilon^2}. \quad (11)$$

其中  $\rho$  為誤差函數， $\alpha$  為基礎高頻權重， $\beta$  控制紋理顯著區域的加權強度。

#### 4.3.2 訓練與測試流程

於訓練階段，模型使用 Training set 之原始影像與其對應二值分割遮罩作為輸入，學習在指定的前景區域內生成符合真實影像分佈之肺結核病灶，並與背景影像保持結構一致性。完成訓練後，模型將能根據輸入之真實背景影像與指定的前景區域遮罩，生成具備高紋理真實性之肺結核 CT 影像。

於測試階段，模型以 Test set 之真實背景影像與其對應的前景區域遮罩作為輸入，其中真實影像的前景區域事先進行遮蔽 (masking) 處理。模型需於該區域內生成相應的肺結核病灶 CT 影像。生成結果將與對應之真實影像進行比較，並透過 4.3.3 所述之評估指標進行定量分析，以評估模型在紋理真實性與細節還原能力上的表現。

最終，為驗證模型之臨床可用性，本計畫擬選取乾淨無結節之肺部 CT 影像作為背景，並以真實病灶之前景遮罩作為條件，生成一組肺結核病灶 CT 影像。該組生成影像除了進行定量評估外，亦將邀請經驗豐富之放射科醫師進行圖靈測試，以評估生成影像在臨床判讀上的真實感與可用性。

#### 4.3.3 評估指標

本計畫之目的為生成具備高紋理真實性的肺結核 CT 影像，因此評估指標將聚焦於影像紋理的真實性與細節還原能力。參考 Armi and Fekri-Ershad [24] 以及本實驗室先前關於結節生成的研究 Chen et al. [5]、Huang [4]，本計畫採用以下評估指標。由於本計畫之目標為使用真實的背景影像生成前景區域，因此以下指標皆針對前景區域進行計算。

##### 1. Gabor 紋理特徵距離

本計畫採用 scikit-image 套件 (Van der Walt et al. [25]) 實作 Gabor filter，相關參數設定依據實驗室先前研究之配置 (Huang [4])，其具體數值詳述於第 4.3.1 節。對於每一三維體積影像，本計畫透過 64 組不同頻率與方向之 Gabor 濾波器提取前景區域之紋理回應，並將其轉換為一個 64 維 Gabor 紋理特徵向量，以表徵影像之多尺度與多方向空間頻域特性。

計算真實資料組與生成資料組之 Gabor 特徵向量之分布，並使用 Wasserstein distance 衡量兩組資料在 Gabor 紋理特徵空間中的分布差異，最後回報 Wasserstein distance 的平均值與標準差作為評估指標。

2. 灰階共生矩陣 (**Gray-level Co-occurrence Matrix, GLCM**)  
對於兩群資料之 ROI 區域，使用 pyradiomics 套件 (Griethuysen et al. [26]) 計算包含 Autocorrelation、Joint Average、Cluster Prominence 等 24 種 GLCM 紋理特徵，並以 Wasserstein distance 衡量兩群資料在各紋理特徵上的分佈差異，最後回報所有紋理特徵的平均與標準差。
3. 灰階游程矩陣 (**Gray-level Run Length Matrix, GLRLM**)  
對於兩群資料之 ROI 區域，使用 pyradiomics 套件 (Griethuysen et al. [26]) 計算包含 Short Run Emphasis (SRE)、Long Run Emphasis (LRE)、Gray Level Non-Uniformity (GLN) 等 16 種 GLRLM 紋理特徵，並以 Wasserstein distance 衡量兩群資料在各紋理特徵上的分佈差異，最後回報所有紋理特徵的平均與標準差。
4. **Fréchet Inception Distance (FID)**  
使用 Inception-V3 網路提取 ROI 之深度特徵 (Heusel et al. [27])，計算真實組與生成組在該特徵空間之 Fréchet distance 作為整體影像品質指標。
5. 特徵分佈可視化 (**t-Distributed Stochastic Neighbor Embedding, t-SNE**)  
使用 t-SNE 方法 (Maaten and Hinton [28]) 將真實組與生成組的 FID 特徵降維至二維空間，並繪製散佈圖以視覺化兩組資料在紋理特徵空間中的分佈情形，輔助評估生成影像的紋理真實性。
6. 醫師圖靈測試：本計畫擬於取得較佳紋理分析指標後，邀請經驗豐富之放射科醫師協助進行圖靈測試，以避免定量指標無法完全反映生成影像之臨床可判讀性與真實感。醫師將被要求在不知情的情況下，對一組包含真實影像與生成影像的資料進行分類，並評估其準確率、敏感度與特異度等，以驗證生成影像在臨床判讀上的可行性與真實性。

## 5 預期結果

本計畫期望能取得比較前期研究中使用之原始 DDPM 模型更優異的紋理分析指標，並在醫師圖靈測試中使醫師正確辨識真偽率  $< 30\%$ ，以驗證所提出之改良方法在提升肺結核病灶 CT 影像紋理真實性與臨床可用性上的有效性。

此外，本計畫亦擬於後期嘗試重現實驗室先前在肺結核病灶生成任務中取得優異結果之 GAN 模型 (Huang [4]、Chen [18])，並將生成結果與本計畫之結果進行比較，以全面評估擴散模型在醫學影像生成任務中的優勢與不足，並為後續研究提供更具體之改進方向。

## 6 需要指導教授指導內容

本研究已完成紋理導向損失函數之設計，惟在時間條件化啓用策略、多損失權重平衡機制，以及潛在空間建模決策上，仍涉及模型收斂穩定性與紋理保真度間之權衡。特別是在多損失並行優化情境下，不同損失項所產生之梯度尺度與方向可能存在競爭或干擾效應，進而影響參數更新軌跡與訓練動力學 (training dynamics)。例如：在 Gabor 紋理損失中，本研究將三個平面 ( $xy$ 、 $xz$ 、 $yz$ ) 之誤差進行平均作為整體監督訊號，若不同平面所對應之梯度方向存在差異或競爭關係，即使某一平面之紋理誤差較大，模型仍無法有效聚焦於特定方向之紋理重建。因此，本研究需進一步與教授討論不同損失項對梯度分佈、收斂速度與生成品質之影響，並在理論與實驗層面評估其對優化動力系統穩定性的作用。

此外，在潛在空間建模與像素空間建模之選擇上，表示轉換亦可能改變梯度傳遞路徑與損失曲面形狀，進而影響收斂速度與最終解之性質。上述關鍵設計涉及優化幾何與訓練動力學層級之問題，需指導教授協助確認最適化設計方向與實驗驗證流程。

## References

- [1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising diffusion probabilistic models”. In: *Advances in neural information processing systems* 33 (2020), pp. 6840–6851.
- [2] Jie An et al. “Bring Metric Functions into Diffusion Models”. In: *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*. Ed. by Kate Larson. Main Track. International Joint Conferences on Artificial Intelligence Organization, Aug. 2024, pp. 578–586. DOI: 10.24963/ijcai.2024/64. URL: <https://doi.org/10.24963/ijcai.2024/64>.
- [3] Tiankai Hang et al. “Efficient diffusion training via min-snr weighting strategy”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2023, pp. 7441–7451.
- [4] Wei-Chieh Huang. “以生成對抗網路生成電腦斷層掃描三維肺結節樣本：基於 Gabor 函數之紋理相似性量度與模型選擇指標”. MA thesis. 臺北, Taiwan: 國立臺灣大學醫學工程學研究所, 2021. DOI: 10.6342/NTU202104006. URL: <http://tdr.lib.ntu.edu.tw/jspui/handle/123456789/81900>.
- [5] Yi-Chang Chen et al. “Texture-embedded Generative Adversarial Nets for the synthesis of 3D pulmonary nodules computed tomography images”. In: *Expert Systems with Applications* 274 (2025), p. 126860. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2025.126860>. URL: <https://www.sciencedirect.com/science/article/pii/S0957417425004828>.
- [6] Siqi Yuan et al. “A novel loss function to reproduce texture features for deep learning-based MRI-to-CT synthesis”. In: *Medical Physics* 51.4 (2024), pp. 2695–2706.
- [7] Luigi Sigillo, Shengfeng He, and Danilo Comminiello. “Latent Wavelet Diffusion: Enabling 4K Image Synthesis for Free”. In: *arXiv preprint arXiv:2506.00433* (2025).
- [8] Haoshen Wang et al. “3d meddiffusion: A 3d medical diffusion model for controllable and high-quality medical image generation”. In: *arXiv preprint arXiv:2412.13059* (2024).
- [9] Shi Yin et al. “HC<sup>3</sup> L-Diff: Hybrid conditional latent diffusion with high frequency enhancement for CBCT-to-CT synthesis”. In: *arXiv preprint arXiv:2411.01575* (2024).
- [10] Mingfeng Jiang et al. “Frequency-Aware Diffusion Model for Multi-Modal MRI Image Synthesis”. In: *Journal of Imaging* 11.5 (2025), p. 152.
- [11] Michael Mathieu, Camille Couprie, and Yann LeCun. “Deep multi-scale video prediction beyond mean square error”. In: *arXiv preprint arXiv:1511.05440* (2015).
- [12] Yochai Blau and Tomer Michaeli. “The perception-distortion tradeoff”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 6228–6237.
- [13] Nasim Rahaman et al. “On the spectral bias of neural networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 5301–5310.
- [14] Xiaojiao Xiao, Qinmin Vivian Hu, and Guanghui Wang. “Pyramid Hierarchical Masked Diffusion Model for Imaging Synthesis”. In: *2025 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2025, pp. 1–8.
- [15] Zehong Ma, Ruihan Xu, and Shiliang Zhang. “PixelGen: Pixel Diffusion Beats Latent Diffusion with Perceptual Loss”. In: (2025). arXiv: 2602.02493 [cs.CV]. URL: <https://arxiv.org/abs/2602.02493>.
- [16] Samuel G. Armato et al. “The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans”. In: *Medical Physics* 38.2 (2011), pp. 915–931. DOI: 10.1118/1.3528204.

- [17] National Cancer Institute. *LIDC-IDRI: The Lung Image Database Consortium Image Collection*. The Cancer Imaging Archive. 2015. URL: <https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI>.
- [18] Leng-Jung Chen. “電腦斷層掃描肺腫瘤良惡性判別之深度學習影像特徵擷取”. Masters thesis. 臺北市, 臺灣: 國立臺灣大學醫學工程學研究所, 2018. DOI: 10.6342/NTU201804220. URL: <http://tdr.lib.ntu.edu.tw/jspui/handle/123456789/71850>.
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [20] Alexander Quinn Nichol and Prafulla Dhariwal. “Improved denoising diffusion probabilistic models”. In: *International conference on machine learning*. PMLR. 2021, pp. 8162–8171.
- [21] Ethan Perez et al. “Film: Visual reasoning with a general conditioning layer”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. 1. 2018.
- [22] Alex Kendall, Yarin Gal, and Roberto Cipolla. “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7482–7491.
- [23] B.S. Manjunath and W.Y. Ma. “Texture features for browsing and retrieval of image data”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18.8 (1996), pp. 837–842. DOI: 10.1109/34.531803.
- [24] Laleh Armi and Shervan Fekri-Ershad. “Texture image analysis and texture classification methods-A review”. In: *arXiv preprint arXiv:1904.06554* (2019).
- [25] Stefan Van der Walt et al. “scikit-image: image processing in Python”. In: *PeerJ* 2 (2014), e453.
- [26] Joost J.M. van Griethuysen et al. “Computational Radiomics System to Decode the Radiographic Phenotype”. In: *Cancer Research* 77.21 (Oct. 2017), e104–e107. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-17-0339. eprint: <https://aacrjournals.org/cancerres/article-pdf/77/21/e104/2934659/e104.pdf>. URL: <https://doi.org/10.1158/0008-5472.CAN-17-0339>.
- [27] Martin Heusel et al. “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium”. In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf).
- [28] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. URL: <http://jmlr.org/papers/v9/vandermaaten08a.html>.