Masked Co-attentional Transformer reconstructs 100x ultra-fast/low-dose whole-body PET from longitudinal images and anatomically guided MRI

Yan-Ran (Joyce) Wang1,2,†*, Liangqiong Qu1,†, Natasha Diba Sheyban1,†, Xiaolong Luo5, Jiangshan Wang5, Kristina Elizabeth Hawk1,3, Ashok Joseph Theruvath2, Sergios Gatidis4, Xuerong Xiao1, Allison Pribnow6, Daniel Rubin1,6,‡, and Heike E. Daldrup-Link1,6,‡

1Department of Biomedical Data Science, Stanford University, Stanford, California 94304, USA
2Department of Radiology, Stanford University, Stanford, California 94304, USA
3Department of Nuclear Medicine, Stanford University, Stanford, California 94304, USA
4Department of Diagnostic and Interventional Radiology, University Hospital Tuebingen, 72076, Germany
5School of Engineering, University of Science and Technology of China, 230001, China
6Department of Pediatrics, Pediatric Oncology, Lucile Packard Children’s Hospital, Stanford University, Stanford, California 94304, USA

*Corresponding Author: Yan-Ran (Joyce) Wang, PhD, Department of Biomedical Data Science, Stanford University, 725 Welch Road, Stanford, CA 94304, USA; Email: wangyanran100@gmail.com; phone: (224) 999-3092

†Equal Contribution

‡Senior Authors

Keywords deep learning; transformer; whole-body PET imaging; longitudinal imaging reconstruction
Abstract

Despite its tremendous value for the diagnosis, treatment monitoring and surveillance of children with cancer, whole body staging with positron emission tomography (PET) is time consuming and associated with considerable radiation exposure. 100x (1% of the standard clinical dosage) ultra-low-dose/ultra-fast whole-body PET reconstruction has the potential for cancer imaging with unprecedented speed and improved safety, but it cannot be achieved by the naive use of machine learning techniques. In this study, we utilize the global similarity between baseline and follow-up PET and magnetic resonance (MR) images to develop Masked-LMCTrans, a longitudinal multi-modality co-attentional CNN-Transformer that provides interaction and joint reasoning between serial PET/MRIs of the same patient. We mask the tumor area in the referenced baseline PET and reconstruct the follow-up PET scans. In this manner, Masked-LMCTrans reconstructs 100x almost-zero radio-exposure whole-body PET that was not possible before. The technique also opens a new pathway for longitudinal radiology imaging reconstruction, a significantly under-explored area to date. Our model was trained and tested with Stanford PET/MRI scans of pediatric lymphoma patients and evaluated externally on PET/MRI images from Tübingen University. The high image quality of the reconstructed 100x whole-body PET images resulting from the application of Masked-LMCTrans will substantially advance the development of safer imaging approaches and shorter exam-durations for pediatric patients, as well as expand the possibilities for frequent longitudinal monitoring of these patients by PET.

Introduction

Positron emission tomography (PET) imaging is the gold standard for the diagnosis and treatment monitoring of patients with lymphoma. The metabolic information from 18F-fluorodeoxyglucose [FDG] PET scans provides important information for therapy response assessments in patients with cancer. However, PET images are obtained by injecting patients with a standardized dose of a radiopharmaceutical (e.g., 18F-FDG) and thus inherently require radiation exposure. Exposure to radiation during diagnostic imaging can increase the risk for developing secondary cancer, especially in pediatric populations undergoing multiple serial scans over time. While novel integrated PET/MRI techniques saved ionizing radiation by replacing CT with MRI, whole-body PET/MRI scans can require upwards of 60 minutes of acquisition times, depending on the field of view and patient height. This can cause discomfort and anxiety for patients, and may potentially lead to motion artifacts that degrade image quality. Decreasing the injected radiotracer dose and reducing image data acquisition times are highly desirable advancements. Both actions, however, lower detection of PET annihilation events, which degrade the diagnostic image quality - since image quality is proportional to the number of coincidence events in the PET detector due to radiopharmaceutical positron annihilation. Consequently, reducing radiotracer dose or shortening PET scan durations was limited so far due to negative effects on the image quality.

To address this challenge, convolutional neural networks (CNNs) have been developed, which can augment high-quality PET images from ultra fast or low dose input images. The current algorithms can be split into two categories - U-net-based approaches and generative adversarial network (GAN) - based approaches. While, these studies have demonstrated the capabilities of CNNs to enhance PET images, they suffer from two major limitations: most works focus on (1) a single anatomical region - i.e. brain reconstruction and (2) limited reduction percentage. Most of the existing work has been confined to brain PET reconstruction, whereas whole-body reconstruction is a much more challenging task. Whole-body PET images have higher intra-patient uptake variation (notably, 18F-FDG radiotracer concentration is much higher in the brain and bladder than elsewhere), which can introduce difficulty in reconstructing images. In addition, the limited performance of classical CNN-based algorithms confine dosage reduction percentage of whole-body PET images to half, a quarter, or 6.25% (the lowest dosage achieved so far). While extremely ultra-low-dose PET (1% of the standard dose) is poised to have the greatest clinical impact relative to existing approaches, this approach is burdened by significantly greater noise and artifacts that are difficult for conventional methods to reconcile (Fig. 1b). To the best of our knowledge, there have been no studies investigating 100x whole-body PET reconstruction.

The goal of this work is to develop a deep learning approach that allows for 100x extremely ultra-low-dose whole-body PET reconstruction. To this end, we utilize longitudinal PET/MRI scans that have been predominantly ignored in previous studies. Indeed, a single patient usually undergoes multiple serial PET scans over time for monitoring of disease progression. It is possible that tumor-bearing regions are the only components in the field of view to vary significantly between scans, whereas the rest of the whole-body scans contain a large amount of longitudinally redundant information. In this work, we argue that such redundant information - the similarity between the baseline and follow-up PET/MRI scans for the same patient - can be learned by the proposed model and leveraged to save 18F-FDG injection and scan time. This provides a novel approach to reducing dosage that in turn opens doors for longitudinal radiology imaging reconstruction. To date, few efforts have been reported in this direction. A significant limitation stems from the fact that most conventional methods fail to effectively capture the dynamic information embedded within longitudinal medical scans. A
framework that integrates complex global longitudinal dependency from sequential data with the precise localization from each input stream is crucial for jointly reasoning multi-serial images.

In response to the aforementioned challenge, we propose Masked-LMCTrans, a longitudinal multi-modality co-attentional CNN-Transformer for PET reconstruction. Masked-LMCTrans reconstructs 100x extremely ultra-low-dose follow-up PET by taking advantage of the simultaneously acquired PET/MRI and referring to the paired half-dose baseline PET/MRI with the tumor area masked out – since tumor-bearing regions vary between scans over the course of longitudinal follow-up (Fig. 1a). This is achieved by introducing separate encoding streams for the baseline and follow-up scans that communicate through the co-attentional transformer layers. The co-attentional transformer layers produce attention-pooled features for each stream conditioned on the other and provide extensive interaction between longitudinal scans. Consequently, the correlation and similarity between the baseline and the follow-up scans are completely encapsulated to aid 100x follow-up PET reconstruction. We demonstrate that this joint structure with longitudinal information outperforms a single-stream unified model which only takes the to-be-reconstructed follow-up 100x PET/MRI as inputs in our experiments (Fig. 4). Additionally, we designed the false focal loss (FF loss) to reduce the false positive errors (Fig. 3). We show that the reconstructed PET scans resemble the standard-dose PET scans, an outcome that effectively propels us towards making the radiation dose of serial scans as low as reasonably achievable (the ALARA principle). Masked-LMCTrans is simple to modify for other diagnostic radiology imaging modalities and could potentially serve as a common foundation, thus impacting a broader range of longitudinal imaging reconstruction efforts, including CT and contrast-enhanced MRI.

Results

PET/MRI Cohort

PET imaging alone shows relatively low spatial resolution, which hinders accurate quantification of radiotracer uptake within small structures. This has motivated past development of anatomical co-registration of PET with high-resolution structural MR images. Integrated 18F-FDG PET/MRI is a new technology that offers high morphological soft-tissue contrast and metabolic information for treatment monitoring of many cancers like lymphoma. In this study, we include two whole-body PET/MRI cohorts: (1) a primary cohort from Stanford University and (2) a cross-continental validation cohort from Tübingen University to rigorously assess the generalizability of our findings. We herein focus specifically on pediatric populations with lymphoma, as children are more sensitive to radiation effects than adults and therefore, would benefit more from ultra-low-dose scans. Notably, the proposed model is not confined to this scheme and can be generalized across diseases and populations given enough training data. In the two cohorts of our study, we collected both pre-treatment whole-body PET/MRI scans and post-treatment follow-up scans for each patient, resulting in 34 paired scans (21 patients; a pair consists of one baseline scan and one follow-up scan) for the Stanford cohort and 10 paired scans (10 patients) for the Tübingen cohort. The reduced dosage PET images were generated by unlisting the PET list-mode data and reconstructed based on the percentage of used counts, simulating ultra-low-dose scheme. To the best of our knowledge, large pooled PET/MRI databases containing PET list-mode data which can be used to generate simulated low-dose PET data for AI model training do not exist so far. As such, our collected cohort is unique in that it is among the first longitudinal list mode PET/MRI databases for Artificial Intelligence-enabled studies.

Masked-LMCTrans approach

Our objective was to reconstruct 100x extremely ultra-low-dose follow-up PET images by utilizing both the simultaneously acquired MRI and the referenced baseline PET/MRI scan. The challenge therein was to effectively integrate four multi-serial multi-modality inputs. Inspired by ViLBERT’s success in modeling visual-linguistic representations, we introduced the co-attentional transformer block to provide interaction between longitudinal medical scans. Our model, known as longitudinal multi-modality co-attentional CNN-Transformer (LMCTrans), is illustrated in Fig 1a. Following the classic approach, the 3D whole-body volume was inferred in a slice-by-slice fashion and the predicted 2D slices were stacked together to reconstruct the final 3D prediction. LMCTrans consists of two parallel streams for baseline and follow-up scan processing that communicate through the co-attentional transformer layers. This joint model accommodates the unique processing needs of each input modality: PET and MRI: by taking advantage of the precise localization from the DenseNet encoder; this approach also provides effective information exchange between the baseline and follow-up input streams by using the global self-attention mechanism introduced by the co-attentional transformer block. To avoid introduction of erroneous upstaging for resolved tumors with low metabolic activity in the post-treatment follow-up PET scans, we masked out the tumor area of the baseline PET and thereby named the approach Masked-LMCTrans. Below we detail the essential building blocks of Masked-LMCTrans.

DenseNet encoder for feature extraction. Feature extraction is the most crucial part for the reconstruction model with combined inputs, as it generates the informative set of features and distinct patterns for each input modality. In this study,
we applied DenseNet block as our encoder of choice as it demonstrates strong representational power with convolution operations and dense collectivities (Fig. 1a). The $l^{th}$ layer receives the “collective knowledge” - the feature-maps from all preceding layers, $x_0, x_1, ..., x_{l-1}$ - as inputs:

$$x_l = H_l([x_0, x_1, ..., x_{l-1}])$$

where $[x_0, x_1, ..., x_{l-1}]$ refers to the concatenation of the feature-maps produced in layer 0, 1, ..., $l - 1$. $H_l(\cdot)$ denotes the composite function of three operations: batch normalization (BN), rectified linear unit (ReLU), and 3x3 convolutions (BN-ReLU-Conv). Such attributes introduce diversified features with all complexity levels and richer patterns that benefit the final PET reconstruction. Moreover, PET and MRI need different levels of feature encoding due to their inherent complexity and the initial level of abstraction of their input representations. Therefore, we set PET feature encoder simpler than MRI (one versus two BN-ReLU-Conv, Fig. 1a) before fusing them together, as PET features are themselves already the output of the PET reconstruction network. We further examined the representational power of different CNN encoders (details of DenseNet’s superiority are shown in the following section).

**Co-attentional transformer layer for longitudinal fusion.** While initial feature extractions are essential, equally important is how the baseline PET/MRI and the follow-up PET/MRI relate to one another - e.g. how to make use of the higher quality baseline PET to benefit the reconstruction of the extremely ultra-low-dose follow-up PET. To this end, we propose the use of co-attentional transformer layer to capture the complex long-range temporal dependency. As opposed to CNNs, where the receptive fields are gradually expanded through a series of convolution operations, the self-attention operations inherited in Transformer allow full coverage of the entire input space through token (i.e. region) matching – i.e. each token is “matched” with all tokens within the input. Three vectors are calculated for each token – query, key and value. The token matching computes a dot product score between the query (the token in consideration) and the key (the token being matched with) to weight the value (of the token being matched with). This score determines how much focus to place on other regions of the input as we encode the region at this certain position. In this way, global self-attention - the dependencies between regions even when they are distant - is obtained. For co-attentional transformer layer in Masked-LMCTrans, given intermediate feature representations of baseline PET/MRI and follow-up PET/MRI, this module computes query, key, and value matrices as in a standard transformer block. However, the keys and values from each sequence are passed as inputs to the other sequence’s multi-headed attention block. Consequently, the attention block produces attention-pooled features for each sequence conditioned on the other - in effect performing baseline conditioned follow-up attention in the baseline stream and the follow-up conditioned baseline attention in the follow-up stream. In this manner, when the model processes each region in the to-be-reconstructed follow-up PET/MRI, the module allows it to look at other positions in the referenced baseline PET/MRI for clues that can help lead to a better encoding for this region. As a result, the correlation and similarity between the baseline and the follow-up stream is completely encapsulated. We demonstrated the effectiveness of this module fusing longitudinal PET/MRI in the next section.

**Masked-LMCTrans.** Chemotherapy leads to a decrease in size and metabolic activity of lymphomas. On interim scans, 2-4 weeks after start of chemotherapy, a metabolic tumor response is defined by a decline in tumor metabolic activity at or below the metabolic activity of the liver as an internal standard of reference. Stable disease is defined as unchanged metabolic activity. And progressive disease is defined as increased metabolic activity or development of new tumor nodules. Acknowledging that variance of this nature should not be introduced in the follow-up PET reconstructions, we elected to mask out the prominent tumor regions in the baseline (pre-treatment) PET scans before feeding them into the proposed model (Fig. 1a inputs; Fig. 2b). Given the innate longitudinal dependency in LMCTrans, this simple operation avoids introduction of erroneous upstaging in the follow-up reconstructed PET scans that could be biased if tumors on the referenced baseline scan are not zeroed out (Fig. 2).

Fig. 1b and Fig. 1c show the qualitative and quantitative results of Masked-LMCTrans model. **Masked-LMCTrans** reconstructed PET images demonstrated significantly less noise and much more detailed structure compared to the simulated 1% extremely ultra-low-dose PET. The SSIM (structural similarity index), PSNR (signal-to-noise ratio), and VIF (visual information fidelity) metrics were significantly higher on the Masked-LMCTrans-reconstructed PET. Via pair-wise t-tests ($p < 0.001$), we observed around 15.8% improvement (as opposed to 1% low-dose PET) in SSIM; 23.4% improvement in PSNR; 186% improvement in VIF. These qualitative and quantitative results show that the proposed **Masked-LMCTrans** model effectively reconstructs the extremely ultra-low-dose $^{18}$F-FDG PET scans, and that the overall image quality reached on reconstruction is similar to 100% dose $^{18}$F-FDG PET scans.
**Fig. 1** Proposed Masked-LMCTrans for 1% extremely ultra-low-dose PET/MRI reconstruction. a, The framework of Masked-LMCTrans. The referenced baseline PET (with tumor area masked out as covered in the yellow mask) and MRI, along with the follow-up 1% PET/MRI are fed into the model as combined inputs. The DenseNet feature encoder encodes PET and MRI separately before aggregation, with BN-ReLU-Conv composite operations and dense collectivities. The co-attentional transformer block fuses the information from the baseline and the follow-up (as indicated by the feature-maps colored in orange and blue respectively; the fused feature-maps in the latter layers are mixed colored). The fusion is performed through baseline and follow-up information exchange by query, key, and value (denoted as Q, K, V). In this manner, Masked-LMCTrans reconstructs 1% follow-up PET image making use of the longitudinal similarity. b, Representative post-treatment $^{18}$F-FDG PET/MRI scan of a 14-year-old male patient with Hodgkin lymphoma (HL). The contrast and structural details are significantly improved on Masked-LMCTrans reconstructed PET as opposed to the simulated 1% PET. The red bounding-box shows the spine anatomical structure that is completely missing in the simulated 1% PET, but successfully reconstructed by Masked-LMCTrans, with the help of the referenced baseline PET. The small tumor around the “shoulder” (the arrow points to) in the baseline PET was resolved after treatment and was not shown on the reconstructed PET. c, Evaluation results of the simulated 1% extremely ultra-low-dose $^{18}$F-FDG PET and Masked-LMCTrans reconstructed $^{18}$F-FDG PET scans on the primary Stanford cohort. SSIM: Structural similarity index; PSNR: Signal-to-noise ratio; VIF: Visual information fidelity; All comparisons are to the ground truth standard-count PET images; P-values are calculated using Wilcoxon signed-rank test between the AI-reconstructed PET and the low-count PET.
The loss function is a cornerstone of neural network model and determines the optimizing direction for model training. We applied the commonly used loss functions for the image restoration task – MSE (mean square error) loss and SSIM (structural similarity index measure) loss - and additionally designed the false focal loss (FF loss) for this study. FF loss is specifically proposed for 100x extremely ultra-low-dose PET reconstruction, as 1% PET images harbor substantial noise, making the model prone to induction of false positive errors. Erroneous upstaging on interim scans would lead to intensified treatment and potential side effects in the absence of viable tumor. FF loss alleviates the impact of these false upstaging focal areas in the output PET images by penalizing incorrect hyperintense pixels during each step of gradient descent in the training process. We formulate the FF loss as below:

$$\mathcal{L}_{FF} = \mathbb{E}(l_{\text{recon}} | l_{\text{true}})((l_{\text{recon}} - l_{\text{true}}) \times 1_{A_\tau})$$

where $l_{\text{recon}}$ and $l_{\text{true}}$ refer to the reconstructed output and the ground-truth PET, respectively. The indicator function of $A_\tau$, where $A_\tau$ denotes the subset of pixels satisfying $(l_{\text{recon}} - l_{\text{true}}) > \tau$, is defined as:

$$1_{A_\tau}(x) = \begin{cases} 1, & \text{if } x \in A_\tau \\ 0, & \text{if } x \notin A_\tau \end{cases}$$

**Proposed false focal loss**

Fig. 2 | The advantage of masking out the tumor regions in the referenced baseline PET – Masked-LMCTrans. a-f, red rectangles (upper panel) are enlarged crops of images below (yellow rectangles); axial PET scan (lower panel), fused with T1-weighted MRI (middle panel). LMCTrans takes scans from two time points – the referenced baseline (pre-treatment) PET/MRI scan (a or b), and the to-be-reconstructed follow-up 1% PET/MRI scan (c). The tumors on the referenced baseline PET, if not masked out (a), would introduce erroneous upstaging in the reconstructed image (where the yellow arrow points to; e), as a result of the innate longitudinal dependency of LMCTrans. Masking out the prominent tumor region in the baseline PET before feeding it to the model (b) avoids false upstaging for the resolved tumors with low metabolic activity in the post-treatment follow-up PET reconstruction (as shown in f).
Regardless of its simplicity, the FF loss is highly advantageous (Fig. 3). The threshold $\tau$ further introduces flexibility to balance between false positive and false negative error probabilities. In addition, the SSIM loss encourages production of output images that are structurally similar to the target image. Coupled with MSE loss and SSIM loss, the composite loss function (as below) for optimizing \textit{Masked-LMCTrans} encourages the PET reconstruction process to reduce noise, keep textures, and preserve structural details.

$$\min \mathcal{L} = \lambda_m \mathcal{L}_{\text{MSE}} + \lambda_s \mathcal{L}_{\text{SSIM}} + \lambda_f \mathcal{L}_{\text{FF}}$$

\textbf{The benefit of longitudinal PET/MRI reconstruction}

The key innovation of this study is introducing a role for longitudinal PET/MRI scans for the follow-up 100x PET reconstruction. Therefore, we first asked whether longitudinal information improves image reconstruction quality. We compared \textit{Masked-LMCTrans} with a single stream unified model taking only the to-be-reconstructed 100x PET/MRIs as inputs. The single stream model avoids making changes to the \textit{Masked-LMCTrans} architecture, by only replacing the co-attentional transformer layer with a standard transformer and removing the baseline encoding stream. The model is initialized and trained identically to \textit{Masked-LMCTrans} to provide fair comparison. We compare to this baseline to establish the impact of longitudinal PET/MRI reconstruction with \textit{Masked-LMCTrans} (Fig. 4). Not surprisingly, the single stream model fails to effectively capture structural details. This is because 100x PET possesses too much noise and suffers from significant information loss which cannot be recovered without additional inputs. Meanwhile, \textit{Masked-LMCTrans}-reconstructed PET can harvest much more accurate information, due to its access to the high-quality baseline PET/MRI as a reference. Moreover, the model itself has the capacity to grasp the complex global temporal similarities between longitudinal scans. We quantify the model performance by computing the structural similarity index measure (SSIM) and the visual information fidelity (VIF). On all of the 24 paired PET/MRI testing scans, \textit{Masked-LMCTrans} consistently outperformed the single-stream baseline in SSIM and VIF. We performed one-tailed, paired t-tests and obtained p-values smaller than 0.05 for all subjects, confirming that improvements were statistically significant. Particularly in oncology, serial
radiographic scans are a common part of clinical workflows, and our findings on longitudinal radiology imaging reconstruction offer a timely new perspective and pathway for general dosage reductions within and beyond PET imaging.

**The advantage of co-attentional transformer layer**

Next, we examined the advantages of Masked-LMCTrans processing longitudinal PET/MRIs over networks with pure convolution operations, i.e. the strong baseline U-net. The U-net model is the state-of-the-art in many biomedical imaging applications, including PET reconstruction. In a recent work\textsuperscript{12}, the authors showed its effectiveness in reconstructing 25% reduced-dose PET/CTs. Here, we extended the U-net model to the 100x extremely ultra-low-dose PET/MRI scenario, denoted as U-net-2 (2 stands for the number of input modalities). We further tested out the U-net model with the multi-serial multi-modality PET/MRI inputs, denoted as U-net-4 (baseline half-dose PET, baseline MRI, follow-up 100x PET and follow-up MRI as four combined inputs). This provides a comprehensive understanding regarding the limitation of the current state-of-the-art in addressing 100x PET reconstruction. The qualitative and quantitative results are shown in Fig. 4.

Classic models with pure convolution operations and direct concatenation suffer from a number of drawbacks. First, they cannot extrapolate beyond the limited receptive field. Convolution operations are great at extracting visual features but are not able to modelize the dependencies between them. Second, the immediate local convolution right after the channel-wise concatenation of baseline and follow-up scans sacrifices valuable information that could otherwise be beneficial for 100x PET reconstruction. This was evident from the performance comparison between U-net-4 and U-net-2. Adding baseline PET/MRI, which was proven to be informative according to the above section, U-net-4 still failed to improve the reconstruction qualitatively or quantitatively, resulting in almost the same SSIM, VIF and PSNR as U-net-2. Meanwhile, Masked-LMCTrans - with the help of co-attentional transformer layers - significantly outperformed U-net models. This observation speaks to the tremendous advantage offered by the proposed model and the effectiveness of co-attentional transformer layers in longitudinal feature aggregation.

**Generalizability in an independent external validation**

To rigorously assess the generalizability of the proposed model and validate our findings, we tested out the trained model (based on the Stanford PET/MRI cohort) directly on the Tübingen database. The Tübingen database is an independent cross-continental validation cohort consisting of 10 paired baseline and follow up PET/MRI scans from pediatric lymphoma patients. The qualitative and quantitative reconstruction results on the Tübingen cohort are shown in Fig. 4a, Supplementary Fig. 1, and Supplementary Table 1. The image quality was significantly improved on LMCTrans-reconstructed PET as opposed to the original 100x extremely ultra-low-dose PET images, by 5.15 dB in PSNR, and 4.50% in SSIM, and 0.167 in VIF; taken together, this demonstrated good model generalization across different institutions. This is particularly noteworthy as the patients were scanned at different institutions on different scanners that used distinct image-reconstruction software; none of these variables were used to train the Masked-LMCTrans model. In addition, we compared Masked-LMCTrans to the single-stream transformer model, which only takes the to-be-reconstructed follow-up 100x PET/MRI as inputs, as well as the classic U-net models. On all of the 10 paired PET/MRI scans, Masked-LMCTrans consistently outperformed the single stream transformer baseline and U-net models in SSIM metric; this showcases its consistent ability to enhance the 100x extremely ultra-low-dose PET images through use of longitudinal information in a generalizable manner, that was also robust with respect to variations in scanner hardware and software.
Fig. 4 | The qualitative and quantitative results of Masked-LMCTrans with comparison models – Single-stream Transformer, U-net-4, and U-net-2. a, The histograms of model comparison on the primary Stanford cohort and on the external Tübingen validation cohort. b, Representative post-treatment $^{18}$F-FDG PET/MRI scan of a 14-year-old male patient with Hodgkin lymphoma (HL) from Stanford cohort. The spine structure is completely missed by the other models, but recovered by Masked-LMCTrans, though not perfect as the 1% PET is extremely noisy. c, Representative post-treatment $^{18}$F-FDG PET/MRI scan of a 20-year-old female patient with Diffuse large B cell lymphoma (DLBCL) from Stanford cohort. Masked-LMCTrans enables detailed shape reconstruction in the brain area (captured by the bounding box). d, the same patient as c. Single-stream Transformer, U-net-4, and U-net-2 fail to reconstruct the liver area without introducing erroneous upstaging. While Masked-LMCTrans successfully recovered the liver area. The $SU_{\text{max}}$ of liver was measured by placing a three-dimensional volume of interest over liver (as the yellow circle indicates in 100% standard-dose PET). The $SU_{\text{max}}$ of liver are measured for the same region over the six PET images and listed below each PET image. e, Representative post-treatment $^{18}$F-FDG PET/MRI scan of a 19-year-old female patient with HL from Stanford cohort. Tumor-to-liver $SU_{\text{max}}$ contrast is preserved in all reconstructions. The reconstructions for the unresolved tumor (as the yellow circle indicates in 100% standard-dose PET) from models of transformer family (Masked-LMCTrans, and Single-stream Trans) resembles that of 100% PET in terms of structural fidelity.
Assessment of four CNN feature encoders

As a final step, we examined the different CNN encoders’ impact on feature representation, focusing on standard off-the-shelf encoders: DenseNet\textsuperscript{[9]}, VggNet\textsuperscript{[6]}, ResNet\textsuperscript{[6]}, and EDSR\textsuperscript{[52]}. To be specific, we replaced the initial feature extraction part of LMCTrans with each of the four encoders. Only the layers before the 1\textsuperscript{st} down-sampling of the original architecture were adopted to simplify the feature extraction. Supplementary Fig. 2 illustrates the backbone and the comparison results for the four encoders. When we moved from EDSR (the least performer) to DenseNet (the best performer) with 32 growth rate and 3 blocks, we observed a 2.5 dB increase in PSNR, indicating the superiority of DenseNet in modeling PET/MRI modalities. A reasonable explanation is that in DenseNet, each layer gets direct inputs from previous layers and this extreme reuse of residuals creates deep supervision from the loss function, minimizing gradient-vanishing concerns. Additionally, the dense connectivity pattern between feature maps enables use of diversified features, especially the low-level features that are crucial for the PET reconstruction task. This comprehensive evaluation of state-of-the-art convolutional encoders offers insights for choosing effective local feature representation techniques among various categories for the AI-enabled medical image reconstruction task.

Discussion

In this work, we propose a novel deep learning model to reconstruct 100x (1% of the standard clinical dosage) ultra-low-dose/ultra-fast whole-body PET. To our knowledge, this reconstruction approach – 1% extreme-low-count PET reconstruction – has not been realized to date. Our approach is unique in that it utilizes the paired baseline PET/MRI scan to aid the reconstruction of extremely ultra-low-dose follow-up PET scans along with the high-resolution structural MRI; this is enabled by the proposed model, Masked-LMCTrans – a longitudinal multi-modality co-attentional CNN-Transformer for PET reconstruction wherein the tumor region in baseline PET is masked out. Masked-LMCTrans accommodates the unique processing needs of each input modality by taking advantage of the strong representational power from the DenseNet encoder. Meanwhile, this joint model integrates complex long-term dependency for the longitudinal images by using the global self-attention mechanism introduced by the co-attentional transformer block. The co-attentional transformer block enables follow-up-attended baseline features to be incorporated into follow-up representations (and vice versa). In this manner, Masked-LMCTrans effectively captures the dynamic information from longitudinal PET/MRIs to enable 100x follow-up PET reconstruction.

Our model holds the potential for marked clinical impact, particularly for pediatric patients or those in other high-risk categories, for whom this approach could enable high-quality imaging with lower radiation exposure in the setting of repeat imaging. Moreover, the method demonstrated has the potential impact of being able to deliver excellent care to a higher volume of patients with the help of reduced PET image acquisition time. It also has the potential economic impact of overall decrease in healthcare spending, as a result of reduced injected radiotracer dose.

Through our findings, we would like to motivate the exploration of longitudinal radiographic image reconstruction and enable a new pathway for general dosage reduction in modalities beyond PET. The inability of the conventional methods to allow for jointly reasoning multi-modal images hinders current exploration in this direction, despite the fact that it is common clinical practice for patients to undergo multiple serial scans. We anticipate that the proposed Masked-LMCTrans may serve as a common foundation for a wide array of radiology imaging reconstruction tasks, including those for CT and contrast-enhanced MRI. In addition, the Tübingen external validation dataset demonstrates the generalizability of Masked-LMCTrans across different scanner models and manufacturers. Therefore, the proposed deep learning model holds the promise of advancing the development of safe imaging tests, shortening exam durations, and opening up greater possibilities for frequent follow-up examinations.

There are some limitations of our study. Lymphoma typically has very robust metabolic activity. Further work is needed to assess if the proposed approach can also perform well with less robust and more subtle hypermetabolic lesions. It is necessary to see if it performs equally well in other tissues besides soft tissue and lymph nodes, such as bone. Furthermore, the innate data demand of deep learning models and the lack of large-scale PET/MRI images represent a major bottleneck for further scale-up and for transitioning our research outcomes to practical use. Finally, the performance of the deep learning model for other types of image modalities was not considered in this work. Future work will be necessary to address these challenges.
Methods

Patients. In this retrospective study, Health Insurance Portability and Accountability (HIPAA)-compliant clinical trial data were collated from two participating centers: University of Tübingen, Germany and Stanford University, CA, USA obtained approval from their institutional review boards (IRB). In addition, Stanford University obtained IRB approval to collect de-identified imaging studies in a centralized image registry, along with relevant clinical information (patient age, sex, tumor type). Written informed consent was obtained from all adult patients and parents of pediatric patients. In addition, children were asked to give their assent. Inclusion criteria were comprised of the following: (1) age < 30 years, (2) histologically proven lymphoma and (3) PET/MRI scan at baseline before chemotherapy. Exclusion criteria were (1) MR-incompatible metal implants, (2) claustrophobia, and (3) pregnancy. Between July 2015 and June 2019, we enrolled 23 children and young adults (14 female, 9 male) with lymphoma and a mean age (standard deviation; range) of 17 years (7; range: 6-30 years). Tumor histology consisted of 14 patients with Hodgkin lymphoma, 6 with non-Hodgkin lymphoma and 3 patients with posttransplant lymphoproliferative disorder (PTLD). For Tübingen, 10 patients were enrolled (5 female, 5 male) with a mean age (standard deviation; range) of 14 years (5; range: 3-18 years) and the following distribution of tumor histology: 8 with Hodgkin lymphoma, 2 with non-Hodgkin lymphoma.

Image acquisition. Stanford patients underwent a whole body integrated $^{18}$F-FDG PET/MRI scan at baseline on a 3T Signa PET/MRI scanner (GE Healthcare, Milwaukee, WI, USA), using a 32-channel torso phased array coil and an eight-channel, receive-only head coil. Before the scan, patients had to fast for at least 4 hours and blood glucose levels had to be below 140mg/dl. $^{18}$F-FDG was administered intravenously 60 minutes before the scan at a dose of 3 megabecquerel per kg body weight. The imaging protocol consisted of an axial T1-weighted two-point Dixon Liver Acquisition with Volume Acquisition (LAVA) sequence (repetition time (TR) 4.2 ms, echo time (TE) 1.1, 2.3 ms, flip angle (FA) 5, slice thickness (SL) 5.2 mm) for attenuation correction and a higher-resolution LAVA sequence (TR 4.2 ms, TE 1.7, 3.4 ms, FA 15, SL 3.4 mm) for anatomical co-registration. PET data were acquired simultaneously with MRI scans, using a 25 cm transaxial FOV and 3:30 minute acquisitions per PET bed. PET data was reconstructed using scanner-specific algorithms, (3D OSEM: 28 subsets, 2 iterations, with time of flight and point spread function information), accounting for attenuation from coils and patient cradle.

Tübingen patients underwent a whole body integrated $^{18}$F-FDG PET/MRI scan at baseline on a 3T Signa PET/MRI scanner (Siemens Healthineers, Erlangen, Germany), using a 16-channel torso phased array coil and a 16-channel head coil. Before the scan, patients had to fast for at least 4 hours and blood glucose levels had to be below 140mg/dl. $^{18}$F-FDG was administered intravenously 60 minutes before the scan at a dose of 3 megabecquerel per kg body weight. The imaging protocol consisted of an axial T1-weighted two-point Dixon Volume Interpolated Breathhold Acquisition (VIBE) sequence (TR 3.95 ms, TEs 1.23, 2.46 ms, FA 10°, SL 3 mm) for attenuation correction and anatomical co-registration. PET data were acquired simultaneously with MRI scans, using a 25 cm transaxial FOV and 4 minutes acquisitions per PET bed. PET data was reconstructed using scanner-specific algorithms, (3D OSEM: 21 subsets, 2 iterations), accounting for attenuation from coils and patient cradle.

Radiotracer input data were used to generate $^{18}$F-FDG PET images. Low-dose PET images were simulated by unlisting the PET list-mode data and reconstructing them based on the percentage of used counts. For the Stanford cohort, the list-mode PET input data collected over a time period of 3:30 and 2 seconds were used to simulate 100% and 1% $^{18}$F-FDG dose levels. For data acquired at the Tübingen site, PET Acquisition time was 4 minutes per bed position and low-dose PET images were simulated using the same relative dose levels accordingly.

Image-preprocessing. The pre-processing pipeline aimed to remove the additional burden of the network learning methods to find patterns between scans for final reconstruction. As opposed to traditional single time scan analyses, registration is essential for longitudinal studies to reduce the spatial discrepancies between scans acquired at different times for the same individual. Across all subjects, the follow-up scans were registered to the baseline MRI as the template using affine transformation. We adopted ANTs$^{43}$, which is considered a state-of-the-art medical image registration toolkit. This ensured all of the scans were registered within each patient’s history. In addition, we used ITK-Snap$^{44}$ to label the tumor regions in the baseline scan. The top five most prominent tumors (largest lesions) for each individual were delineated with ellipsoid-shaped masks. These tumor masks were used to mask out the tumor area of the baseline PET images to avoid introducing erroneous upstaging signals for the follow-up reconstruction. Top 0.1% of the pixels in PET images were clipped. Of note, the clipping operation is critical in model convergence and stabilizing training as these top pixels possess extreme high values and are outliers of the distribution. Lastly, all scans were normalized between zero and one before feeding them to the deep learning model.

Data augmentation. Model performance improved with increasing training data sample size. We used random rotation,
random shifting, and random zoom for data augmentation. During each step of stochastic gradient descent in the training process, we perturbed each training sample (both baseline and follow-up PET/MRI images; four combined inputs) with a random rotation between -20 to 20 degrees and with a random shift between -20% to 20% across x and y axis, and with a random zoom between 0.8 to 1.2. Data augmentation resulted in improvement for all models; around 1% improvement in SSIM metric for Masked-LMCTrans and slight improvement for U-net.

**Pretraining on Masked-LMCTrans.** Transformer models yield significant improvements on a wide spectrum of high-level computer vision tasks. Researchers attribute their success partly to the pre-training on large-scale dataset. While, few efforts are made to explore the role of pretraining in transformer models for low-level vision tasks. To the best of our knowledge, the two pioneers exploring this point are IPT\(^{45}\) and EDT\(^{46}\) in 2021. Moreover, the effect of pre-training in medical image reconstruction tasks has never been explored. More detailed analysis is in real need to understand how pre-training affects the representations of biomedical neural network models. In our experiment, we initialized Masked-LMCTrans with the weights of ViT-base\(^{47}\), pretrained on ImageNet21k\(^{48}\) and ImageNet2012\(^{49}\). We find that pre-training plays a strikingly important role in 100x low-count PET reconstruction; around 10% improvement in SSIM metric for Masked-LMCTrans. Without pre-training, the proposed model is not able to generate predictions that are a plausible translation of the input 100x extremely ultra-low-dose PET images.

**Model architecture.** Transformer architectures expanded to image processing very recently and soon became the game changer in computer vision. Vision Transformer (ViT), a transformer adapted for image processing, has shown impressive performance on high-level vision tasks\(^{29,30}\), but few efforts are made to explore its role in low-level image reconstruction tasks. In this study, the proposed Masked-LMCTrans model is among the pioneering efforts of utilizing transformer for medical image reconstruction. In terms of the inputs to the model, we adopted 2.5D input scheme to ensure vertical spatial consistency. Five consecutive axial slices from both the baseline and the follow-up PET and MRI modalities are fed into the model as combined inputs, resulting in 20 input slices in total.

**Training details.** Following ResNet\(^{41}\) and ViT\(^{47}\) (vision transformer), we used a learning rate warmup for 5 epochs and then linearly decay the learning rate over the course of training. We trained the models with Adam\(^{50}\) optimizer, using \(\beta_1 = 0.9\) and \(\beta_2 = 0.99\). We adopted three-fold cross-validation for the Stanford cohort. Each fold has 23 paired baseline and follow-up PET/MRI scans (from 14 patients) for training, 8 paired scans for testing, and 3 paired scans for validation (from 7 patients). The training time using four NVIDIA GeForce RTX 3090 GPUs with 24GB VRAM was about 12 hours, and the reference time for each subject was only 10 seconds.

**Computational assessment.** For evaluation, we adopted three quantitative metrics to measure the quality of the reconstructed PET images: SSIM (the structural similarity index), PSNR (peak signal-to-noise ratio), and VIF (Visual information fidelity). SSIM is the most widely used metric in imaging reconstruction. It is a combination of luminance, contrast, and structural similarity comparison functions\(^{31,51}\). Specifically, the SSIM score was derived by comparison of the AI-reconstructed PET and true standard-dose PET sequences and quantified the similarity on a scale of zero (no similarity) to one (perfect similarity). PSNR is most commonly used to measure the reconstruction quality of a lossy transformation\(^{52}\). The higher the PSNR, the better the degraded image has been reconstructed to resemble the original image. SSIM and PSNR mainly focus on pixel-wise similarity; thus, VIF is introduced. VIF uses natural statistics models to evaluate psychovisual features of the human visual system\(^{32}\). The code for calculating the performance was written with Python. Standardized uptake values (SUVs) are the most widely used metric in clinical FDG-PET oncologic imaging in assessing tumor glucose metabolism. The \(\text{SUV}_{\text{max}}\) of liver were measured by placing a three-dimensional volume of interest over liver. The SUV metric in the study was measured using OsiriX version 12.5.1. (OsiriX software). SUV values were calculated based on patient body weight by using the equation: \(\text{SUV} = \frac{\text{tissue tracer activity}}{\text{in millicuries per milliliter}} \times \frac{\text{injected dose (in millicuries)/patient body weight (in grams)}}\).

**Lessons from model training and experiments.** We examined the difference of using slices from the axial plane or coronal plane and found that axial demonstrates superior performance. More details are provided in Supplementary Fig. 3. For the training scheme, we tried out multi-task learning with segmentation added besides the objective reconstruction, but did not notice improvement as shown in other reconstruction-enhanced segmentation studies\(^{54,55}\).
Acknowledgements
This study was supported by a grant from the National Cancer Institute of the US National Institutes of Health, grant number R01CA269231, and the Andrew McDonough B+ Foundation.

Author Contributions
All authors meet the criteria detailed in Author Instructions. Initial study concept and design: Y.J.W., L.Q. Acquisition of data: N.D.S., A.J.T., S.G., and A.P. Model training: Y.J.W, X.L., and J.W. Analysis and interpretation of data: Y.J.W., X.L., and J.W. Primary drafting of paper and figures: Y.J.W. Critical revision of the manuscript for important intellectual content: H.D., N.D.S., and K.E.H. Final approval of the paper: All authors.

Competing Interests
The authors declare no competing interests.

Data Availability
All of the code of the algorithm, training and testing the models is available at GitHub. The imaging data comes from Stanford Healthcare and is not publicly available. The de-identified data is available from the authors upon reasonable request and with permission of the institutional review board.
Baum, S. H. et al. Contribution of PET/CT to prediction of outcome in children and young adults with rhabdomyosarcoma. Journal of Nuclear Medicine 52, 1535-1540 (2011).

Kleis, M. et al. Diagnostic value of PET/CT for the staging and restaging of pediatric tumors. European journal of nuclear medicine and molecular imaging 36, 23-36 (2009).

Baratto, L. et al. PET/MRI Improves Management of Children with Cancer. Journal of Nuclear Medicine 62, 1334-1340 (2021).

Huang, B., Law, M. W.-M. & Khong, P.-L. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. Radiology 251, 166-174 (2009).

Leide-Svegborn, S. Radiation exposure of patients and personnel from a PET/CT procedure with 18F-FDG. Radiation protection dosimetry 139, 208-213 (2010).

Meulepas, J. M. et al. Radiation exposure from pediatric CT scans and subsequent cancer risk in the Netherlands. JNCI: Journal of the National Cancer Institute 111, 256-263 (2019).

Brenner, D. J. et al. Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proceedings of the National Academy of Sciences 100, 13761-13766 (2003).

Mathews, J. D. et al. Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. Bmj 346 (2013).

Tian, J. et al. Does the novel integrated PET/MRI offer the same diagnostic performance as PET/CT for oncological indications? PLoS One 9, e90844 (2014).

Fendler, W. P., Czernin, J., Herrmann, K. & Beyer, T. Variations in PET/MRI operations: results from an international survey among 39 active sites. Journal of Nuclear Medicine 57 (2016).

Jadvar, H. & Colletti, P. M. Competitive advantage of PET/MRI. European journal of radiology 83, 84-94 (2014).

Chaudhari, A. S. et al. Low-count whole-body PET with deep learning in a multicenter and externally validated study. NPJ digital medicine 4, 1-11 (2021).

Theruvath, A. J. et al. Validation of Deep Learning-based Augmentation for Reduced 18F-FDG Dose for PET/MRI in Children and Young Adults with Lymphoma. Radiology: Artificial Intelligence 3, e200232 (2021).

Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, 234-241 (2015).

Goodfellow, I. et al. Generative adversarial nets. Advances in neural information processing systems 27 (2014).

Xiang, L. et al. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing 267, 406-416 (2017).

Xu, J., Gong, E., Pauly, J. & Zaharchuk, G. 200x low-dose PET reconstruction using deep learning. arXiv preprint arXiv:1712.04119 (2017).

Schramm, G. et al. Approximating anatomically-guided PET reconstruction in image space using a convolutional neural network. NeuralImage 224, 117399 (2021).

Chen, K. T. et al. True ultra-low-dose amyloid PET/MRI enhanced with deep learning for clinical interpretation. European Journal of Nuclear Medicine and Molecular Imaging, 1-10 (2021).

Chen, K. T. et al. Generalization of deep learning models for ultra-low-count amyloid PET/MRI using transfer learning. European Journal of Nuclear Medicine and Molecular Imaging 47, 2998-3007 (2020).

Wang, T. et al. Machine learning in quantitative PET: A review of attenuation correction and low-count image reconstruction methods. Physica Medica 76, 294-306 (2020).

Whiteley, W., Luk, W. K. & Gregor, J. DirectPET: full-size neural network PET reconstruction from sinogram data. Journal of Medical Imaging 7, 032503 (2020).

Wang, Y.-R. J. et al. Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure. European Journal of Nuclear Medicine and Molecular Imaging, 1-11 (2021).

Schmalt, J. P., Surti, S., Otero, H. J., Servaes, S. & Karp, J. S. Investigating Low-Dose Image Quality in Whole-Body Pediatric 18F-FDG Scans Using Time-of-Flight PET/MRI. Journal of Nuclear Medicine 62, 123-130 (2021).

Unterrainer, M. et al. Recent advances of PET imaging in clinical radiation oncology. Radiation Oncology 15, 1-15 (2020).

Ellis, S. & Reader, A. J. Longitudinal multi-dataset PET image reconstruction. 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 1-3 (2017).

Ellis, S. & Reader, A. J. Synergistic longitudinal PET image reconstruction. 2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD), 1-3 (2016).
Chen, S., Bortsova, G., García-Uceda Juárez, A., Tulder, G. v. & Bruijne, M. d. Multi-task attention-based semi-supervised learning for medical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 457-465 (2019).

Bahrami, K., Rekik, I., Shi, F. & Shen, D. Joint reconstruction and segmentation of 7T-like MR images from 3T MRI based on cascaded convolutional neural networks. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 764-772 (2017).
Supplementary Figure 1. The model comparison evaluated on the external Tübingen validation set – Masked-LMCTrans, Single-stream Transformer, U-net-4, and U-net-2. Representative post-treatment $^{18}$F-FDG PET/MRI scan of a 14-year-old female patient with Hodgkin lymphoma (HL) from Tübingen cohort. Single-stream Transformer, U-net-4, and U-net-2 fail to reconstruct the liver area without introducing erroneous upstaging or preserve the structure details. While Masked-LMCTrans recovered the liver area. The SUV$_{\text{max}}$ of liver was measured by placing a three-dimensional volume of interest over liver (as the red circle indicates in 100% standard-dose PET). The SUV$_{\text{max}}$ of liver are measured for the same region over the six PET images and the values are listed below each PET image. The spine structure is completely missed by the other three models, but relative clear depicted by Masked-LMCTrans, though not perfect as the 1% PET is extremely noisy. Note that patients from the Tübingen cohort were not used to train the Masked-LMCTrans model, making it a true external validation dataset.
Supplementary Table 1. Performance metrics of Masked-LMCTrans for 1% extremely ultra-low-dose PET/MRI reconstruction evaluated on Tübingen cohort. Evaluation results of the simulated 1% extremely ultra-low-dose $^{18}$F-FDG PET and Masked-LMCTrans reconstructed $^{18}$F- FDG PET scans on the Tübingen cohort. SSIM: Structural similarity index; PSNR: Signal-to-noise ratio; VIF: Visual information fidelity; All comparisons are to the ground truth standard-count PET images; P-values are calculated using Wilcoxon signed-rank test between the AI-reconstructed PET and the low-count PET.

|                  | 100x ultra-low-dose PET (N =10) | Masked-LMCTrans PET (N = 10) | P-value |
|------------------|---------------------------------|------------------------------|---------|
| SSIM             |                                 |                              |         |
| Mean (SD)        | 0.747 (0.045)                   | 0.899 (0.028)                | <0.001  |
| Median (Q1, Q3)  | 0.764 (0.699, 0.779)            | 0.900 (0.872, 0.920)         |         |
| PSNR             |                                 |                              |         |
| Mean (SD)        | 27.4 (0.99)                     | 34.4 (1.61)                  | <0.001  |
| Median (Q1, Q3)  | 29.2 (28.7, 30.3)               | 35.0 (34.0, 35.6)            |         |
| VIF              |                                 |                              |         |
| Mean (SD)        | 0.112 (0.010)                   | 0.254 (0.026)                | <0.001  |
| Median (Q1, Q3)  | 0.113 (0.106, 0.116)            | 0.257 (0.234, 0.270)         |         |
Supplementary Figure 2. CNN encoder architectures and performance comparisons. In this study, we examined the representational power of four most commonly used CNN feature extractors – DenseNet, ResNet, EDSR, and Vgg. We replaced the CNN-encoder part of Masked-LMCTrans with the four structures respectively. The training and testing are performed on only one-fold of the three-fold cross-validations applied, with 23 paired training scans and 8 paired testing scans. a-d, The architecture and operation composition for DenseNet, ResNet, EDSR, and Vgg encoder respectively. e, The histograms show the quantitative metrics in PSNR (the peak signal-to-noise ratio), VIF (the visual information fidelity), and SSIM (the structural similarity index). The DenseNet encoder thanks to the “dense collectivity” attribute presents superior performance in all metrics as opposed to the other CNN encoders.
Supplementary Figure 3. Performance comparison of models with inputs from the axial plane or the coronal plane. The training and testing were performed with 25 training scans and 9 testing scans, on Single-stream Transformer model. The axial view inputs with five consecutive slices outperformed coronal view inputs with five consecutive slices in PSNR (the peak signal-to-noise ratio), VIF (the visual information fidelity), and SSIM (the structural similarity index).