## Supplementary Materials

### Methods

*U-Net Implementation and Training:* The encoder portion of the pre-trained network using Site 1 data is composed of layers which perform 3×3 two-dimensional convolutions, batch normalization, and rectified linear unit activation operations. The dimensionality of the data is reduced by 2-by-2 max pooling. In the decoder portion, the data in the encoder layers are concatenated with those in the decoder layers and restored to its original dimensions by linear interpolation. The network was trained with a learning rate of 0.0002, batch size of four, the L1-norm as the loss function, adaptive moment estimation as the optimization method, and trained for 100 epochs. Since residual learning is used, the output is added to the input low-count image for the final synthesized PET image [17]. This method was also used to train networks in Method C and Method D. For the following studies (investigating the optimal low-count image and investigating the effect of incomplete data) no fine-tuning was performed and Site 2 data were directly used as inputs.

*Investigating the optimal low-count image:* The Site 2 data were framed for 12 seconds, 24 seconds, 1 minute, and 5 minutes from the start of PET acquisition (Site 2) for reconstruction to produce low-count (approximately 1%, 2%, 5%, 25% “dose”) PET images. These images were used as inputs for the low-count PET channel of the pre-trained network. The T1-weighted image was used as the input for the missing T2 FLAIR channel. 8 datasets from Site 1 were set aside as testing data for the network. The low-count image with the closest image quality metrics (peak signal-to-noise ratio: PSNR, structural similarity: SSIM, and root mean square error: RMSE) as the results obtained from the Site 1 data was selected for use for subsequent analyses. Paired t-tests were used to assess the difference of metrics across datasets [28].

*Investigating the effect of incomplete data:* The T1- and T2-weighted images as well as a T1-derived binary head mask and matrices of either all ones or all zeros were used as inputs for the missing T2 FLAIR channel in the network. 5%-dose images were used as the low-count input for Site 2 data. Paired t-tests were used to assess the difference of metrics across datasets [28].

*Selection of the Pre-trained Network:* The training from Site 1 data resulted in five different networks. The five networks were applied on the Site 2 data using Methods A and B. The image quality metrics (PSNR, SSIM, and RMSE) were obtained and compared using the repeated measures ANOVA and subsequently pair-wise paired t-tests, both at the significance level p=0.05/3 (Bonferroni correction).

**Results**

Qualitatively, all synthesized images showed much improvement in noise reduction compared to the low-count image and resemble the ground truth image, except when using zeros to replace the missing contrast, where undesirable errors arose outside the head (Figure S1). Quantitatively, the 5% images used as inputs resulted in the most similar metrics compared to those from site 1 [28]; 5% was thus selected as the dose reduction factor for Site 2 data in this study (Figure S2).

Moreover, replacing the T2 FLAIR data with T1, T2, or head mask images had similar performance for data acquired on Site 2 (Figure S3, $p>0.05$ for most comparisons; $p<0.05$ for comparing SSIM with head mask replacement vs. T1 or T2 replacement) [28], thus the T1 images were arbitrarily selected as the replacement for the main study.

Repeated measures ANOVA and pair-wise paired t-tests have shown at p=0.05 with Bonferroni correction that while all three metrics according to ANOVA and 19/30 of the Method A t-test comparisons were significantly different, much fewer (5/30) of the Method B results are significantly different and the ANOVA results are not significantly different except for PSNR (p=0.01, close to threshold). The results are shown in Table S4. From these results and from our observation that all

Method A results perform worse than the other methods we believe a random selection of one such pre-trained network is suitable for this study.

**Figure S1.** Representative images as well as the difference images for the preliminary studies conducted to choose a low-count ("low-dose") image as well as a replacement for the missing input channel [28].

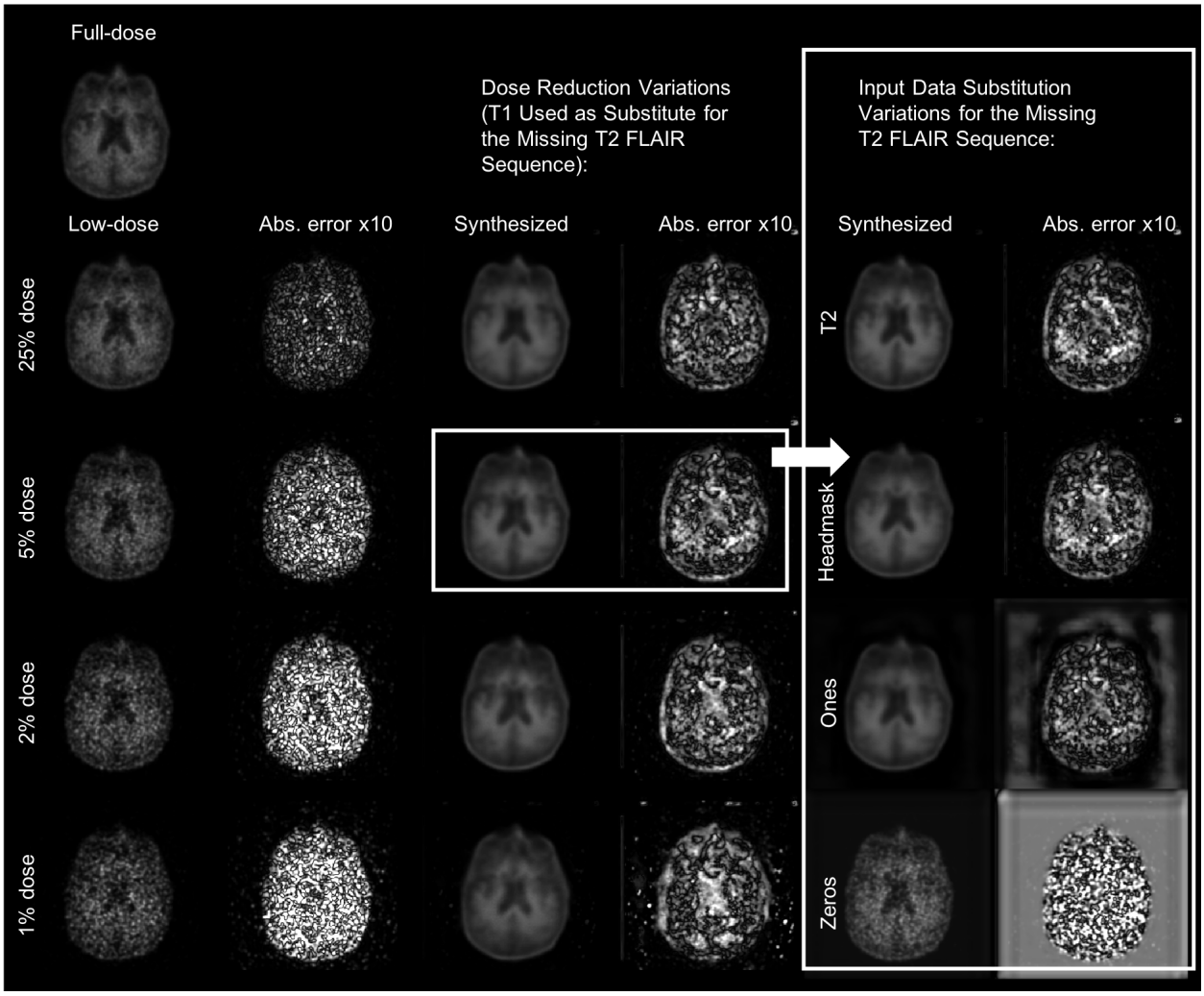

**Figure S2.** Image quality metrics comparing the original low-count ("low-dose") image from Site 2 with its synthesized counterpart from the network [28].

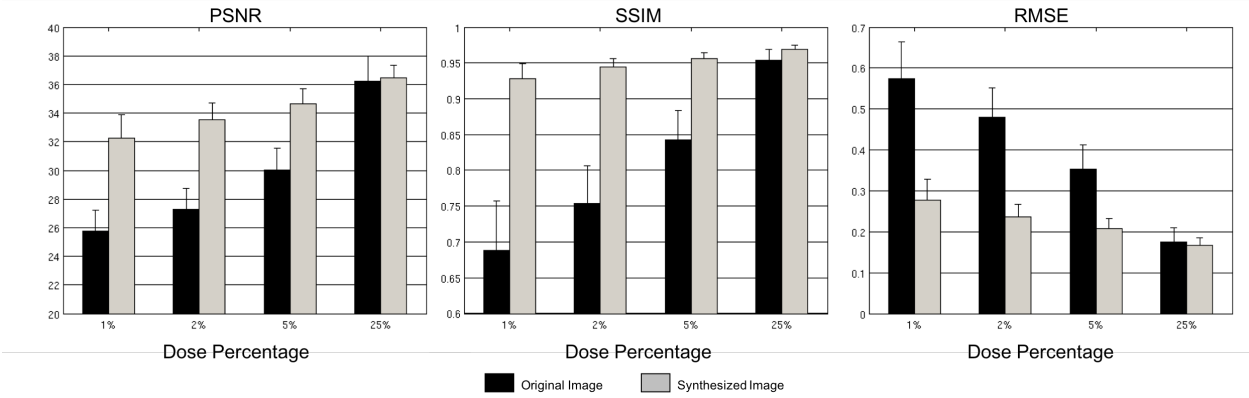


**Figure S3.** Metrics comparing the full-dose images to the low-count ("low-dose") as well as those synthesized with data substitution for the missing T2 FLAIR contrast [28].

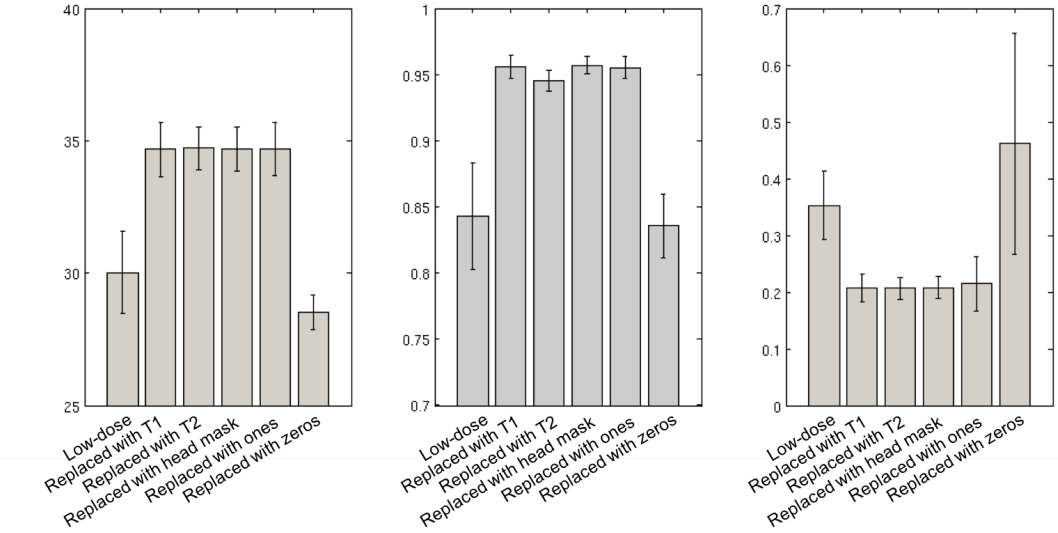

**Table S1**. Acquisition parameters for Site 2 data

| Sequence | $FOV_{read}$ | $FOV_{phase}$ | #Slices | Slice Thickness | $N_{read}$ | $N_{phase}$ | TR [ms] | TE [ms] | FA [deg] |
|---|---|---|---|---|---|---|---|---|---|
| T1-weighted (MPRAGE) | 250 | 250 | 176 | 1 | 512 | 512 | 1900 | 2.53 | 9 |
| T2-weighted (TSE) | 172.5 | 230 | 44 | 3 | 288 | 384 | 6000 | 100 | 120 |

**Table S2**. Confusion matrices for the amyloid status readings between image types. Network training methods: A: direct application of pre-trained network; B: transfer learning starting with pre-trained network; C: training new network from scratch; D: training new network with combined datasets

| | | Images from Method A | | |
|---|---|---|---|---|
| | | Negative | Positive | Total |
| Ground-truth images | Negative | 80 | 4 | 84 |
| | Positive | 30 | 46 | 76 |
| | Total | 110 | 50 | 160 |
| | | Images from Method B | | |
| | | Negative | Positive | Total |
| Ground-truth images | Negative | 78 | 6 | 84 |
| | Positive | 0 | 76 | 76 |
| | Total | 78 | 82 | 160 |
| | | Images from Method C | | |
| | | Negative | Positive | Total |
| Ground-truth images | Negative | 82 | 2 | 84 |
| | Positive | 1 | 75 | 76 |
| | Total | 83 | 77 | 160 |
| | | Images from Method D | | |
| | | Negative | Positive | Total |
| Ground-truth images | Negative | 79 | 5 | 84 |
| | Positive | 1 | 75 | 76 |
| | Total | 80 | 80 | 160 |

**Table S3**. Image quality scores (1: uninterpretable; 2: poor; 3: adequate; 4: good; 5: excellent) from the four readers. Network training methods: A: direct application of pre-trained network; B: transfer learning starting with pre-trained network; C: training new network from scratch; D: training new network with combined datasets

Reader 1

| Image quality score | Ground truth | Method A | Method B | Method C | Method D | Short-time |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 9 | 0 | 0 | 0 | 7 |
| 3 | 3 | 13 | 7 | 0 | 3 | 9 |
| 4 | 3 | 13 | 15 | 21 | 16 | 15 |
| 5 | 34 | 5 | 18 | 19 | 21 | 9 |
| Mean | 4.78 | 3.35 | 4.28 | 4.48 | 4.45 | 3.65 |

Reader 2

| Image quality score | Ground truth | Method A | Method B | Method C | Method D | Short-time |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 18 | 3 | 0 | 2 | 18 |
| 3 | 22 | 16 | 16 | 17 | 14 | 18 |
| 4 | 17 | 6 | 13 | 17 | 22 | 4 |
| 5 | 0 | 0 | 8 | 6 | 2 | 0 |
| Mean | 3.40 | 2.70 | 3.65 | 3.73 | 3.60 | 2.65 |

Reader 3

| Image quality score | Ground truth | Method A | Method B | Method C | Method D | Short-time |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 34 |
| 2 | 11 | 5 | 0 | 3 | 5 | 6 |
| 3 | 21 | 34 | 2 | 12 | 18 | 0 |
| 4 | 8 | 1 | 34 | 24 | 15 | 0 |
| 5 | 0 | 0 | 4 | 1 | 2 | 0 |
| Mean | 2.93 | 2.90 | 4.05 | 3.58 | 3.35 | 1.15 |

Reader 4

| Image quality score | Ground truth | Method A | Method B | Method C | Method D | Short-time |
|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 37 |
| 2 | 25 | 21 | 1 | 1 | 1 | 3 |
| 3 | 15 | 12 | 13 | 7 | 10 | 0 |
| 4 | 0 | 7 | 18 | 20 | 21 | 0 |
| 5 | 0 | 0 | 8 | 12 | 8 | 0 |
| Mean | 2.38 | 2.65 | 3.83 | 4.08 | 3.90 | 1.08 |

**Table S4.** Image metrics results for the use of different pre-trained networks in Methods A and B. P-values lower than the significance threshold (0.05/3) are in bold. The F critical value is calculated at the alternative probability of 5%.

Method A

| PSNR (t-tests, p-value) | Network 2 | Network 3 | Network 4 | Network 5 | PSNR (ANOVA) | |
|---|---|---|---|---|---|---|
| Network 1 | 0.46 | **$<10^{-3}$** | **$<10^{-3}$** | 0.83 | Degrees of Freedom ($df_1$, $df_2$) | 4, 156 |
| Network 2 | N/A | **0.002** | **$<10^{-3}$** | 0.32 | F statistic | 14.14 |
| Network 3 | N/A | N/A | 0.82 | **$<10^{-3}$** | Critical value | 2.43 |
| Network 4 | N/A | N/A | N/A | **$<10^{-3}$** | p-value | **$<10^{-3}$** |

| SSIM (t-tests, p-value) | Network 2 | Network 3 | Network 4 | Network 5 | SSIM (ANOVA) | |
|---|---|---|---|---|---|---|
| Network 1 | 0.70 | **0.004** | **$<10^{-3}$** | 0.50 | Degrees of Freedom ($df_1$, $df_2$) | 4, 156 |
| Network 2 | N/A | **$<10^{-3}$** | **$<10^{-3}$** | 0.71 | F statistic | 24.77 |
| Network 3 | N/A | N/A | **$<10^{-3}$** | **$<10^{-3}$** | Critical value | 2.43 |
| Network 4 | N/A | N/A | N/A | **$<10^{-3}$** | p-value | **$<10^{-3}$** |

| RMSE (t-tests, p-value) | Network 2 | Network 3 | Network 4 | Network 5 | RMSE (ANOVA) | |
|---|---|---|---|---|---|---|
| Network 1 | 0.97 | **0.002** | **$<10^{-3}$** | 0.59 | Degrees of Freedom ($df_1$, $df_2$) | 4, 156 |
| Network 2 | N/A | **0.002** | **$<10^{-3}$** | 0.63 | F statistic | 12.71 |
| Network 3 | N/A | N/A | 0.35 | **$<10^{-3}$** | Critical value | 2.43 |
| Network 4 | N/A | N/A | N/A | **$<10^{-3}$** | p-value | **$<10^{-3}$** |

Method B

| PSNR (t-tests, p-value) | Network 2 | Network 3 | Network 4 | Network 5 | PSNR (ANOVA) | |
|---|---|---|---|---|---|---|
| Network 1 | 0.14 | 0.26 | 0.57 | 0.05 | Degrees of Freedom ($df_1$, $df_2$) | 4, 156 |
| Network 2 | N/A | **0.01** | 0.09 | 0.61 | F statistic | 3.44 |
| Network 3 | N/A | N/A | 0.56 | **0.016** | Critical value | 2.43 |
| Network 4 | N/A | N/A | N/A | **0.01** | p-value | **0.01** |

| SSIM (t-tests, p-value) | Network 2 | Network 3 | Network 4 | Network 5 | SSIM (ANOVA) | |
|---|---|---|---|---|---|---|
| Network 1 | 0.81 | 0.95 | 0.40 | 0.39 | Degrees of Freedom ($df_1$, $df_2$) | 4, 156 |
| Network 2 | N/A | 0.84 | 0.81 | 0.34 | F statistic | 0.54 |
| Network 3 | N/A | N/A | 0.66 | 0.40 | Critical value | 2.43 |
| Network 4 | N/A | N/A | N/A | 0.14 | p-value | 0.71 |

| RMSE (t-tests, p-value) | Network 2 | Network 3 | Network 4 | Network 5 | RMSE (ANOVA) | |
|---|---|---|---|---|---|---|
| Network 1 | 0.21 | 0.28 | 0.61 | 0.06 | Degrees of Freedom ($df_1$, $df_2$) | 4, 156 |
| Network 2 | N/A | 0.04 | 0.15 | 0.59 | F statistic | 2.91 |
| Network 3 | N/A | N/A | 0.60 | **0.016** | Critical value | 2.43 |
| Network 4 | N/A | N/A | N/A | **0.014** | p-value | 0.02 |