DUAL-GLOW: Conditional Flow-Based Generative Model for Modality Transfer
Supplementary Material

In this supplement, We provide extensive additional details regarding the neuroimaging experiments and some theoretical analysis.

Contents

1. The ADNI dataset

Data used in the experiments for this work were obtained directly from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biologicalmarkers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD).For up-to-date information, see `www.adni-info.org`.

Pre-processing details are described in the main paper. MR images were segmented such that the skull and other bone matter were masked, leaving only grey matter, white matter, and cerebrospinal fluid (CSF). After pre-processing, we obtain 806 clean MRI/PET pairs. The demographics of the final dataset are shown in Table 1. As voxel size was fixed for all volumes to $1.5mm^3$, processed images were of dimension $121 \times 145 \times 121$. Images were cropped and downsampled slightly after skull extraction to allow for faster training.

Table 1: Demographic details of the full ADNI dataset in our experiments.

CATEGORY	CN	SMC	EMCI	MCI	LMCI	AD
# of subjects	259	18	88	263	64	114
Age (mean)	76.47	70.94	71.64	77.86	73.14	75.98
Age (std)	5.26	4.76	6.55	7.42	6.43	7.27
Gender (F/M)	116/143	11/7	38/50	66/197	24/40	42/72

Table 2: Main hyperparameters in our natural image experiments.

DATASET	Levels	Depth	Layers (Relation Nets)
UT-Zap50K	5	8	8
Cartoon-Celeba	6	8	8

2. Architecture Details

2.1. ADNI Brain Imaging Experiments

There are 4 "levels" in our two invertible networks, each "level" containing 16 affine coupling layers. The small network with three 3D convolutional layers is shared by two nonlinear operators $s(\cdot)$ and $t(\cdot)$. There are 512 channels in the intermediate layers. For the hierarchical correction learning network, we split the hidden codes of the output of the first three modules in the invertible network and design four 3D CNNs (relation networks) with 1 convolutional layer for all latent codes. For the conditional framework case, we concatenate the five discriminators with the adaptive number of layers to the tail of all four levels of the MRI inference network and the top-most level of the PET inference network.

For other compasion methods, we have 13 resBlocks of 34 *convs* with the U-net architecture for C-VAE, 12 *convs* in *G* and 8 resBlocks in *D* for the cGAN, 15 *convs* in *G* and 8 resBlocks in *D* for UcGAN. Flow-based methods for image translation do not currently exist. But we implemented an iterative Glow (iGlow: 4 levels, 4×10 coupling layers), concatenating paired MRI and PET as input. After training, we fix networks and set input PET as trainable variables and obtain PET by iteratively optimizing the log-likelihood w.r.t. these variables. But the unstable gradient ascent gives bad results.

2.2. Natural Image Experiments.

For natural image experiments, the settings of the hypersparameters are shown in Table 2. The depth is equal to the number of coupling layers in each "level". Since the resolution of images in UT-Zap50K is 128×128, we use the "level" of 5. For the celebA dataset (resolution: 256×256), the number of the "level" is 6. The relation network is a CNNs with 8 conv layers in both two experiments.

3. Generated Samples

3.1. Brain Imaging Experiments

The images that follow are additional representative samples from our framework. Figures 1, 2, 3 show ground truth and reconstructions of cognitively healthy, mildly cognitively impaired, and Alzheimer's diseased patients within our test group.

Figure 4 shows the comparison visualization results, which shows that DUAL-GLOW outperforms other methods in most regions.

Figure 5 shows additional age conditioning results, again for each of the three disease groups (CN, MCI, and AD). We also plot the mean intensity of other 30 ROIs of 3 subjects given 6 age labels (from 50 to 100) in Fig 6, which shows a clear decreasing trend, i.e., decreased metabolism with aging.

3.2. Other Experiments

While not the focus of our work, we show some additional results on two standard imaging datasets: the cartoon-to-face translation in Figure 7; the sketch-to-shoe snynthesis in Figure 8. The input of the CelebA face dataset is the cartoon image processed by using the technique in *opencv*. For UT-Zap50K shoe dataset, we extract edge images as sketches by using HED (Holistically-nested edge detection) and learn a mapping from sketch to shoe.

4. Theoretical Analysis

4.1. Hierarchical architecture

In Fig 3/paper, half of the feature map is used as input to the next level. The computational complexity is mainly dependent on the input of each coupling layer. Let ℓ denote the number of levels. Supposing the input size is 2^ℓ and the time complexity for each level is $\mathcal{O}(N)$, we have the time complexity of $\mathcal{O}(2^\ell N\ell)$ for the flat architecture and $\mathcal{O}((2^{\ell+1}-2)N)$ for the hierarchical one. A larger ℓ leads to further reduction.

4.2. The choice of λ

λ regularizes networks for MRI. Our goal is to model the conditional distribution rather than the joint distribution. A lower weight on this constraint (0.001 in all experiments) leads to easier optimization and better qualitative results (Fig 9).

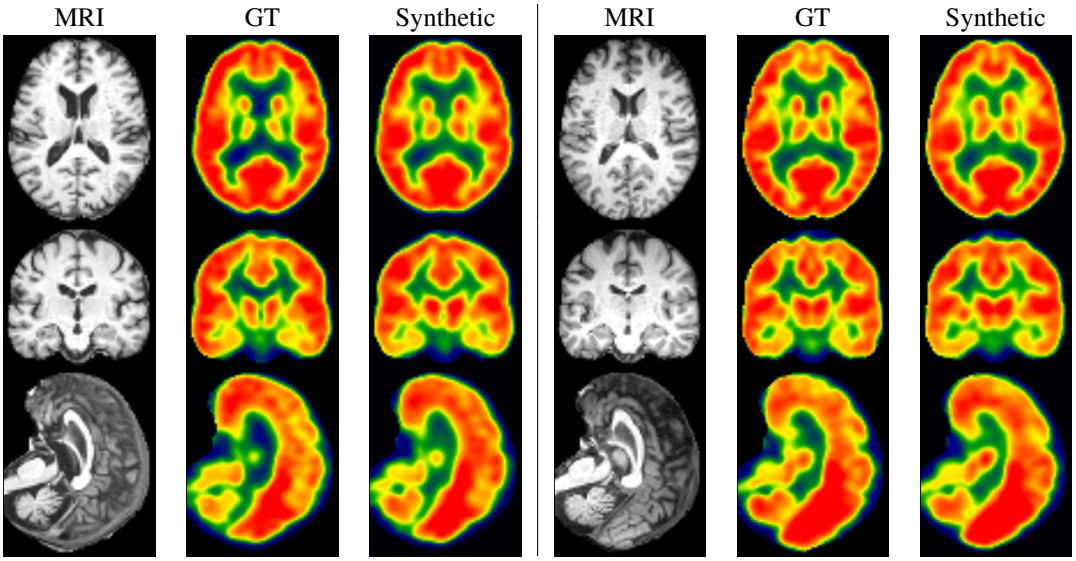

Figure 1: Two synthetic CN subjects. From left to right: input of MRI, the ground truth, the synthetic subject.

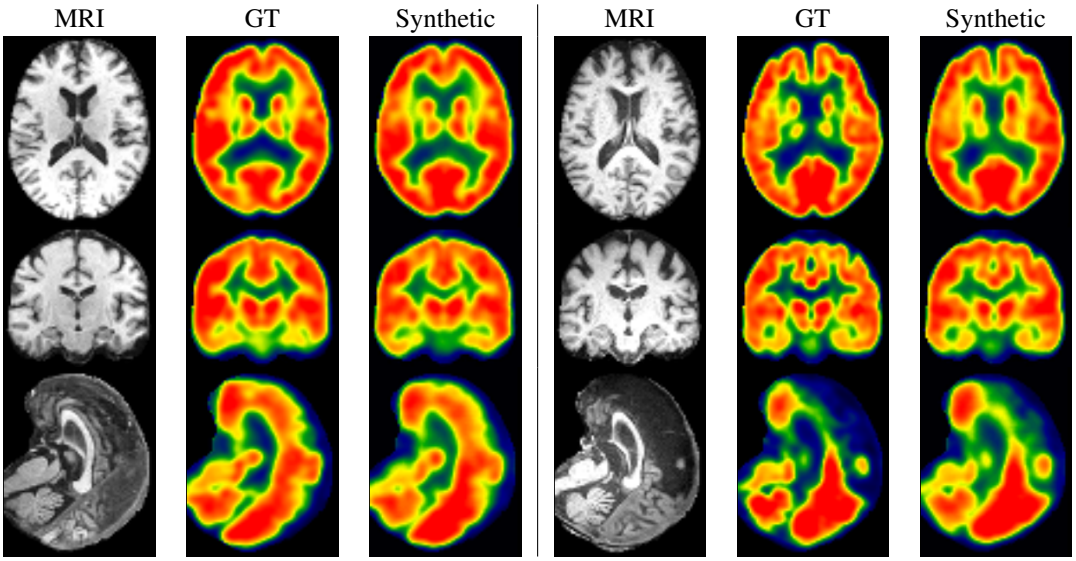

Figure 2: Two synthetic MCI subjects. From left to right: input of MRI, the ground truth, the synthetic subject.

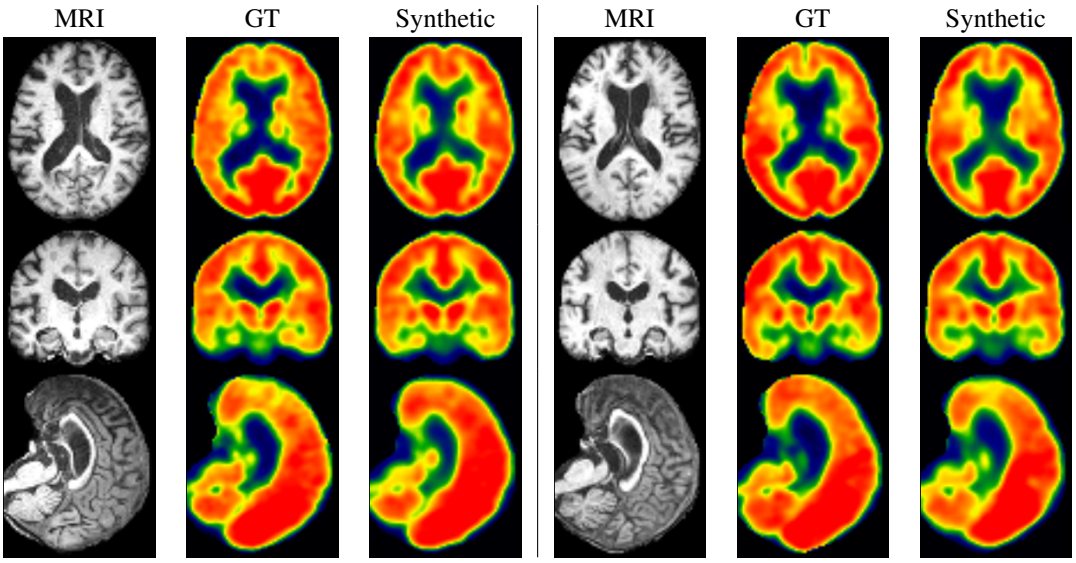

Figure 3: Two synthetic AD subjects. From left to right: input of MRI, the ground truth, the synthetic subject.

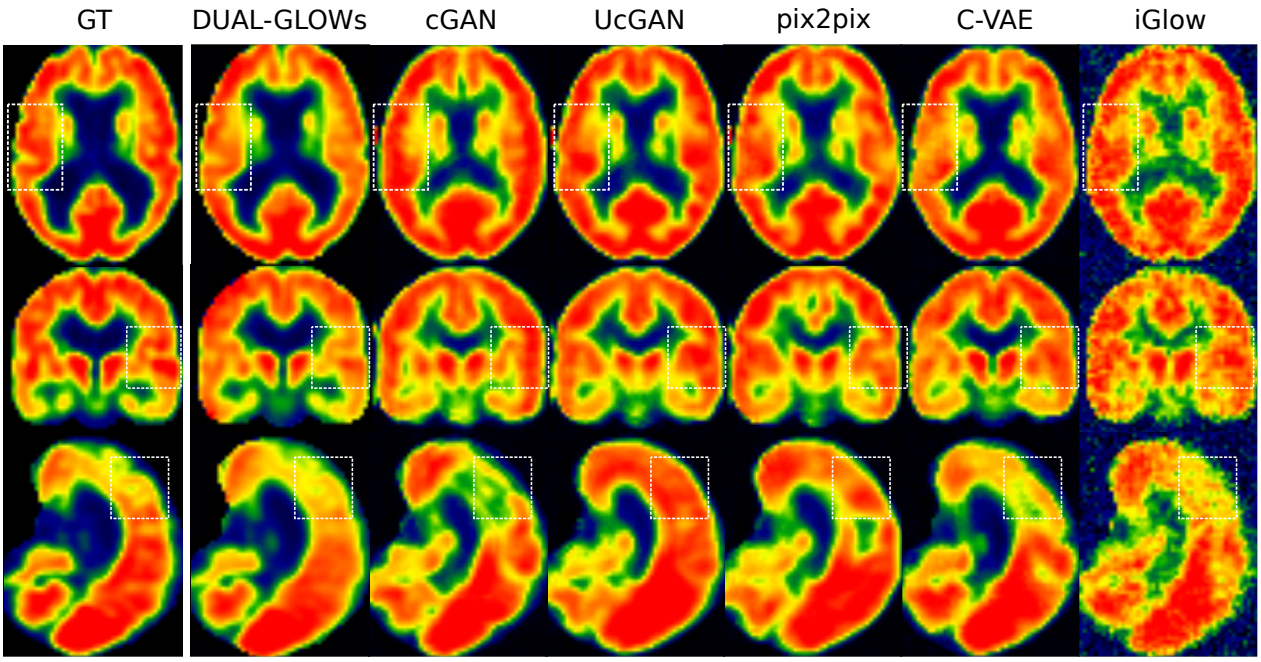

Figure 4: DUAL-GLOW can produce more accurate prediction in dashed rectangles.

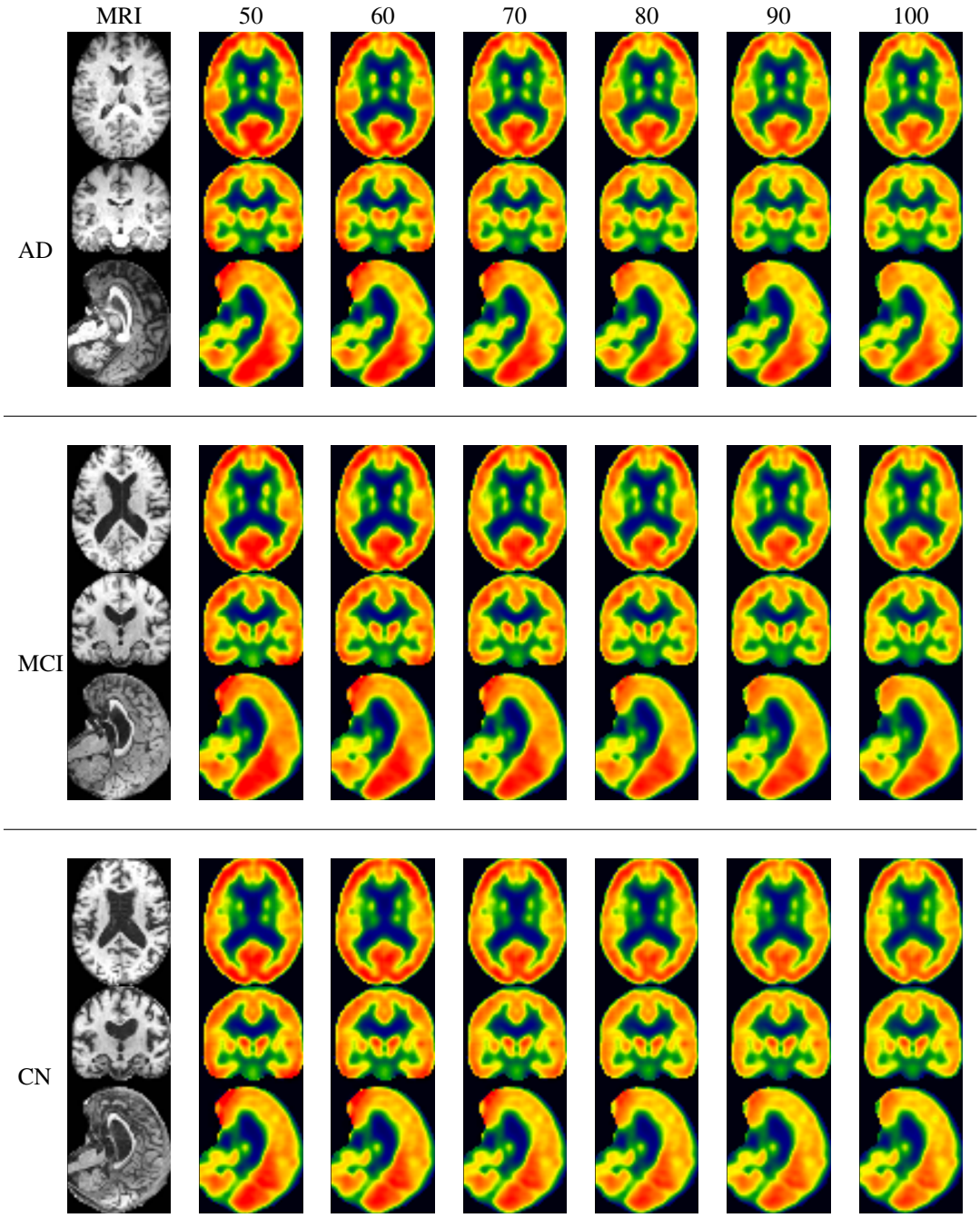

Figure 5: Age information manipulation. There are 3 subjects, AD, MCI, and CN, each subject provides 6 results *w.r.t.* a variant of age labels. As we scan left to right, we indeed see a decrease trend in metabolism (less red, more yellow) which is completely consistent with what we would expect in aging.

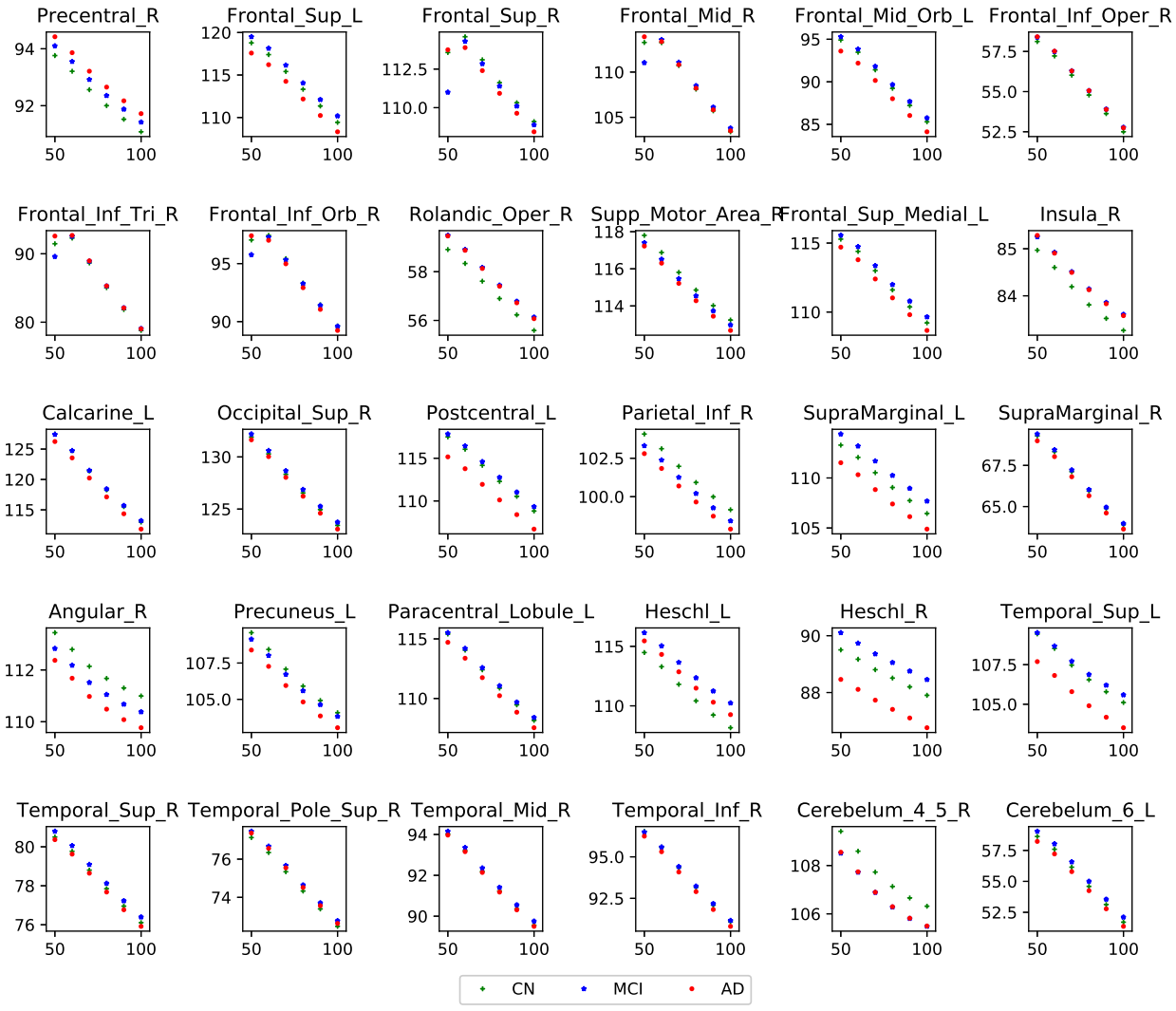

Figure 6: Decreasing trends for 30 ROIs (related to aging).

Figure 7: Input cartoon images and generated/reconstructed faces applying our DUAL-GLOW framework to the CelebA dataset.

Figure 8: Input "sketch" images and generated/reconstructed shoes applying our DUAL-GLOW framework to the UT-Zap50K dataset.

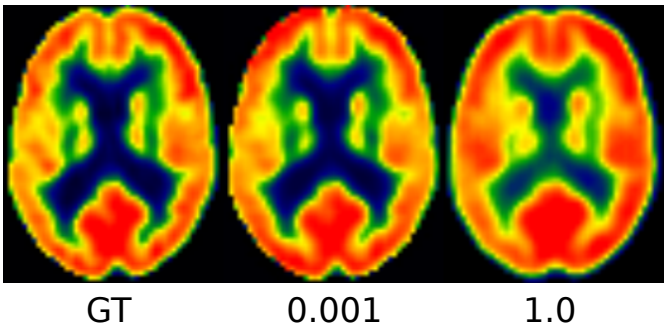

Figure 9: Better for the small λ.