Supplementary: On Statistical Analysis of Neuroimages with Imperfect Registration

Won Hwa Kim Sathya N. Ravi Sterling C. Johnson Ozioma C. Okonkwo Vikas Singh

University of Wisconsin, Madison, WI

http://pages.cs.wisc.edu/∼wonhwa

1. Demonstration of Group Analysis using Synthetic FDG-PET Images

In this supplement, we demonstrate statistical group analysis (diseased versus control) pipeline on a population of synthetically generated 2-D FDG-PET scans where the images are *not* brought into perfect alignment prior to analysis. We seek to demonstrate that our algorithm is able to still pick up the synthetically introduced group-level differences. We compare the results from the standard approach and our framework. Separately, we provide Matlab code so that an interested reader can run a simple "demo" version of the codebase, to better understand the behavior of the method.

1.1. Dataset

Using two template images μ_{control} and μ_{disease} as representatives for the two groups, 40 images (20 diseased and 20 control images) are generated by applying a random affine transformation (with rotation r and translation t) and adding noise of $N(0, 0.3)$, as shown in Fig. 1. Results of this procedure are demonstrated in Fig. 2 using certain random translations in x and y directions as well as a random rotation. In this case, visual inspection suggests that the groups vary in the "red" region (note that individual participants are iid draws from the disease and control distributions).

$$I = A_{r,t}\mu_{\text{group}} + N(0, 0.3) = \begin{cases} A_{r,t}\mu_{\text{control}} + N(0, 0.3) & \text{if control} \\ A_{r,t}\mu_{\text{disease}} + N(0, 0.3) & \text{if diseased} \end{cases}$$

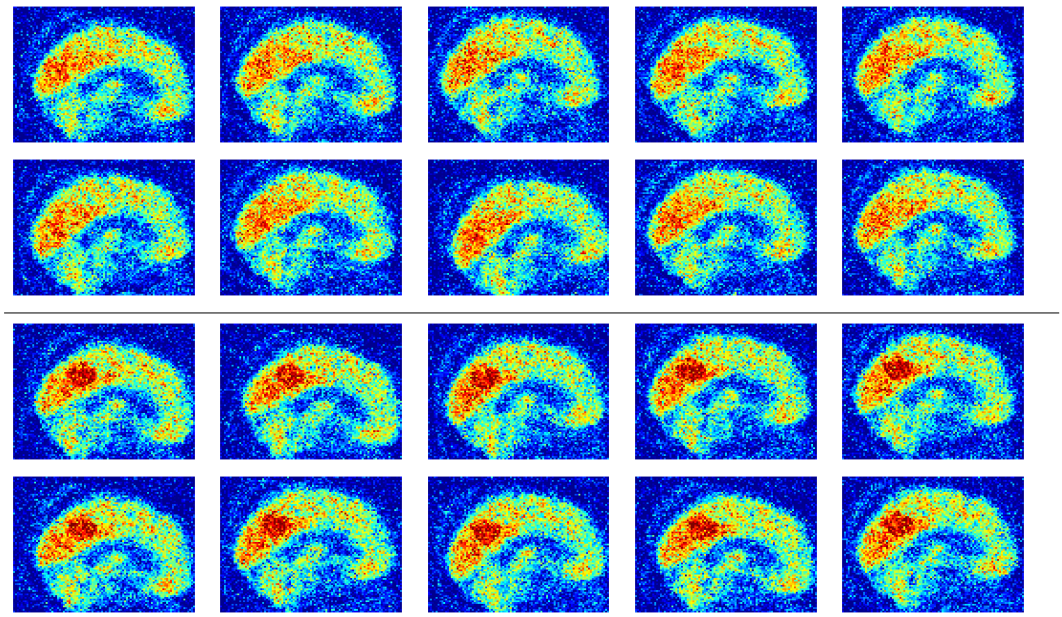


Figure 1: First 10 images from each group (top: control, bottom: diseased). The images are rotated and translated randomly. The visual inspection shows that the groups vary in the cuneus region (in red).

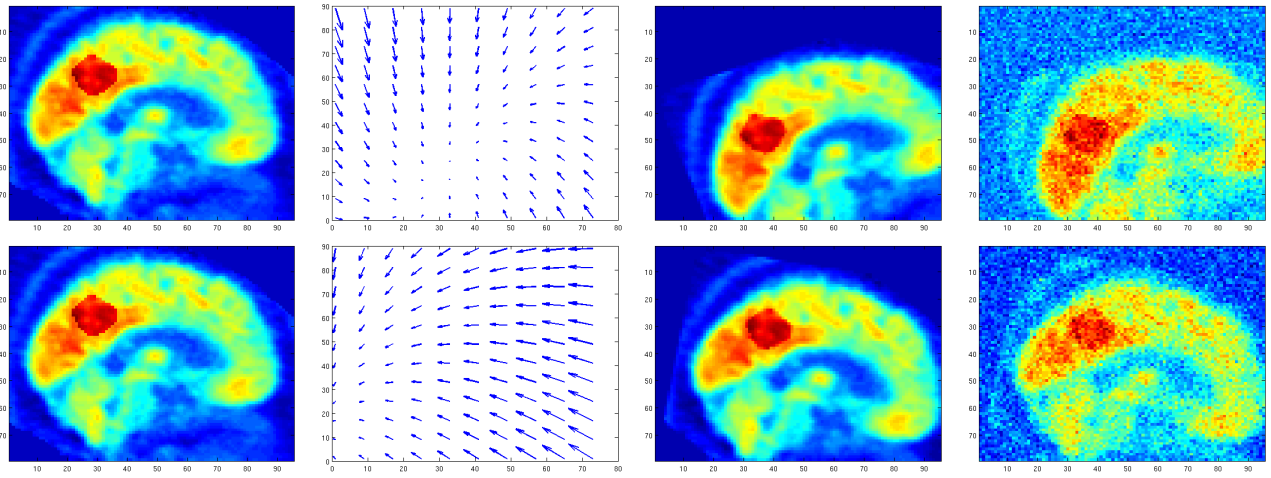


Figure 2: Data generation process. First column: template image, Second column: deformation field, Third column: warped image using the deformation field, Fourth column: warped image with noise.

1.2. Standard Statistical Group Analysis Framework

In a typical statistical group analysis, we assume that there exist two groups of samples in the dataset and information about which sample is which group is known (class labels). The analysis proceeds by performing a hypothesis test (e.g., Student's t-test) at each pixel location, comparing the empirical distributions of the image intensities from the two groups. The Null hypothesis (H_0) is that the distributions are not statistically different. If the disease affects a specific pixel, we should be able to reject the Null with high statistical confidence. We therefore obtain a pixel-wise p-value.

In general, we reject the null hypothesis if the p-value is lower than $\alpha = 0.05$ level, meaning that there indeed exist group difference at that pixel position. An example of the two distributions from two different groups is shown in Fig. 3. But since we are performing a large number of such tests at each pixel, we must perform multiple hypothesis testing correction to ensure that the false positives are not reported. This controls the family-wise error rate. Applying Bonferroni correction, we set the threshold α/n instead of α, where α is the threshold level and n is the number of tests. The most important assumption in this framework is that the images are registered perfectly, meaning that pixel-to-pixel correspondence is given as shown at the top row of Fig. 4. However, when the registration is not perfect, we end up with errors in the transformation and lose the pixel-to-pixel correspondence as shown at the bottom row of Fig. 4.

To summarize, the standard group analysis pipeline on images is as follows:

1. Apply hypothesis testing at each pixel position.
2. Apply multiple comparisons correction.
3. Display the resultant p-value map in the original space.

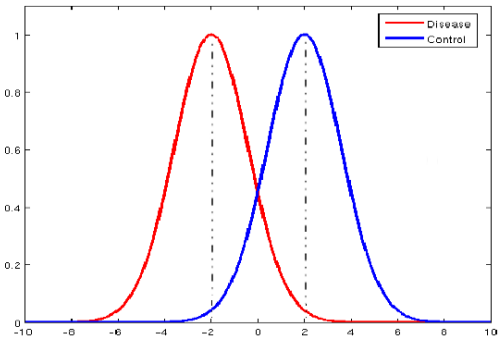


Figure 3: Two distributions from two different groups.

1.3. Statistical Group Analysis Result

We first demonstrate the result using the standard statistical group analysis pipeline. The result in Fig. 5 shows that as the perturbation level increases (variance in rotation and translation in the transformation), this pipeline fails to detect the true group level difference. In contrast, applying our framework to the same data with 5% level perturbation, we are able to identify some regions that are similar to the ground truth. Our framework is as follows:

1. Construct graph structure from the images.
2. Apply scattering transform to obtain scattering coefficients.
3. Apply hypothesis testing on the scattering coefficients at each pixel location.

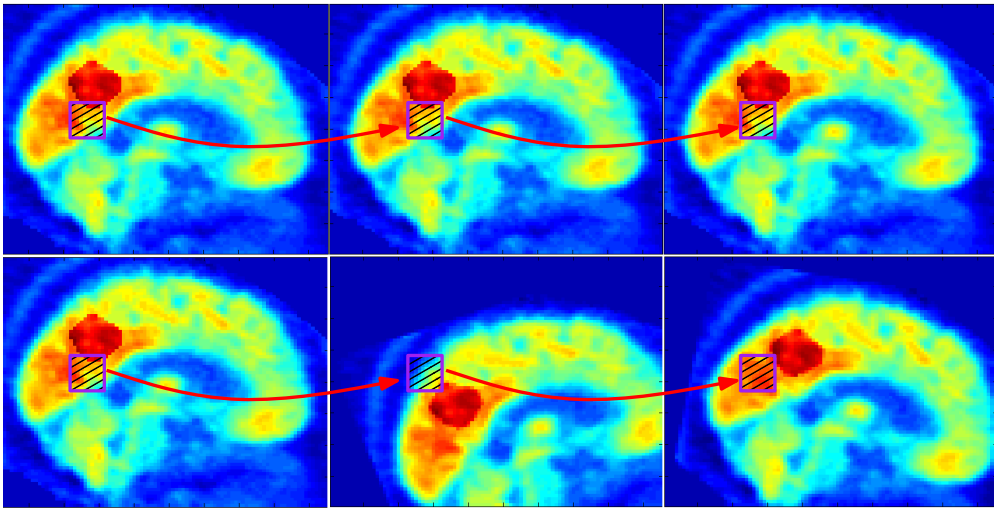
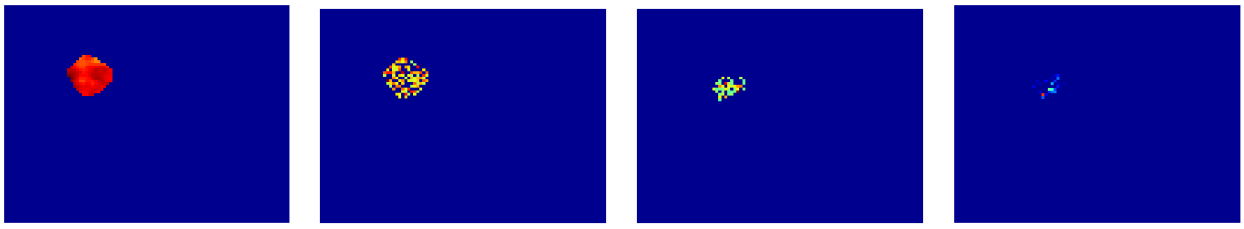


Figure 4: Top: Pixel-to-pixel correspondences. Top: Correct matches with proper registration, Bottom: False matches due to errors in the registration process. Notice that the follow-up statistical test will have poor statistical power because the distributions being evaluated are not representative.

4. Apply multiple comparisons correction.

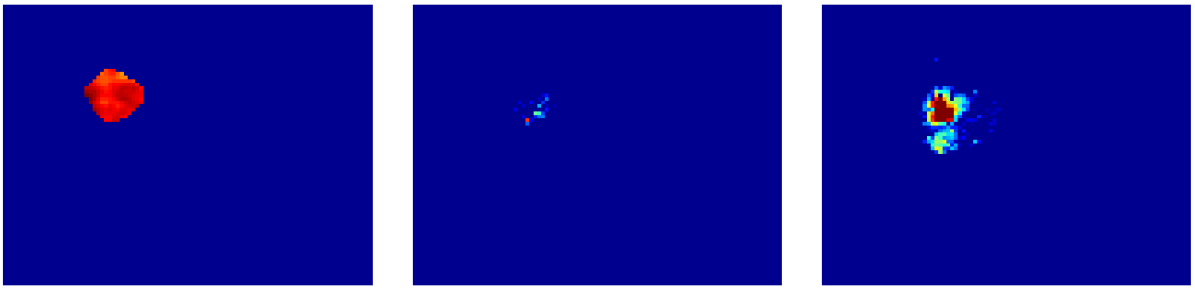
5. Display the resultant p-value map in the original space.

As seen in Fig. 6, even at 5% level perturbation where the standard approach fails to detect any group difference, our framework identifies the group differences with very low p-values. This shows that even when there exists errors in the registration process and traditional approach does not work, our algorithm will be able to correctly identify the group differences that are statistically significant.



(a) (b) (c) (d)

Figure 5: Comparison of resultant p-value maps from group analysis using the standard analysis with different levels of deformation. a) ground truth, b) no perturbation, c) 3% perturbation, d) 5% perturbation. As we increase the perturbation level, the statistical power to detect the true group differences in (a) decreases.



(a) (b) (c)

Figure 6: Comparison of resultant p-value maps from group analysis on the raw images. a) ground truth, b) standard analysis on 5% perturbed inputs, c) analysis from our algorithm on the same 5% perturbed inputs. Despite the deformations, our algorithm detects the group differences.