The 5th annual Symposium on Data Science and Statistics
CS20 - Neural Network Analysis

# A Novel Network Architecture Combining Central-Peripheral Deviation with CNNs for DTI Studies

Soyun Park, University at Buffalo
June 9, 2022

# Contents

# Contents

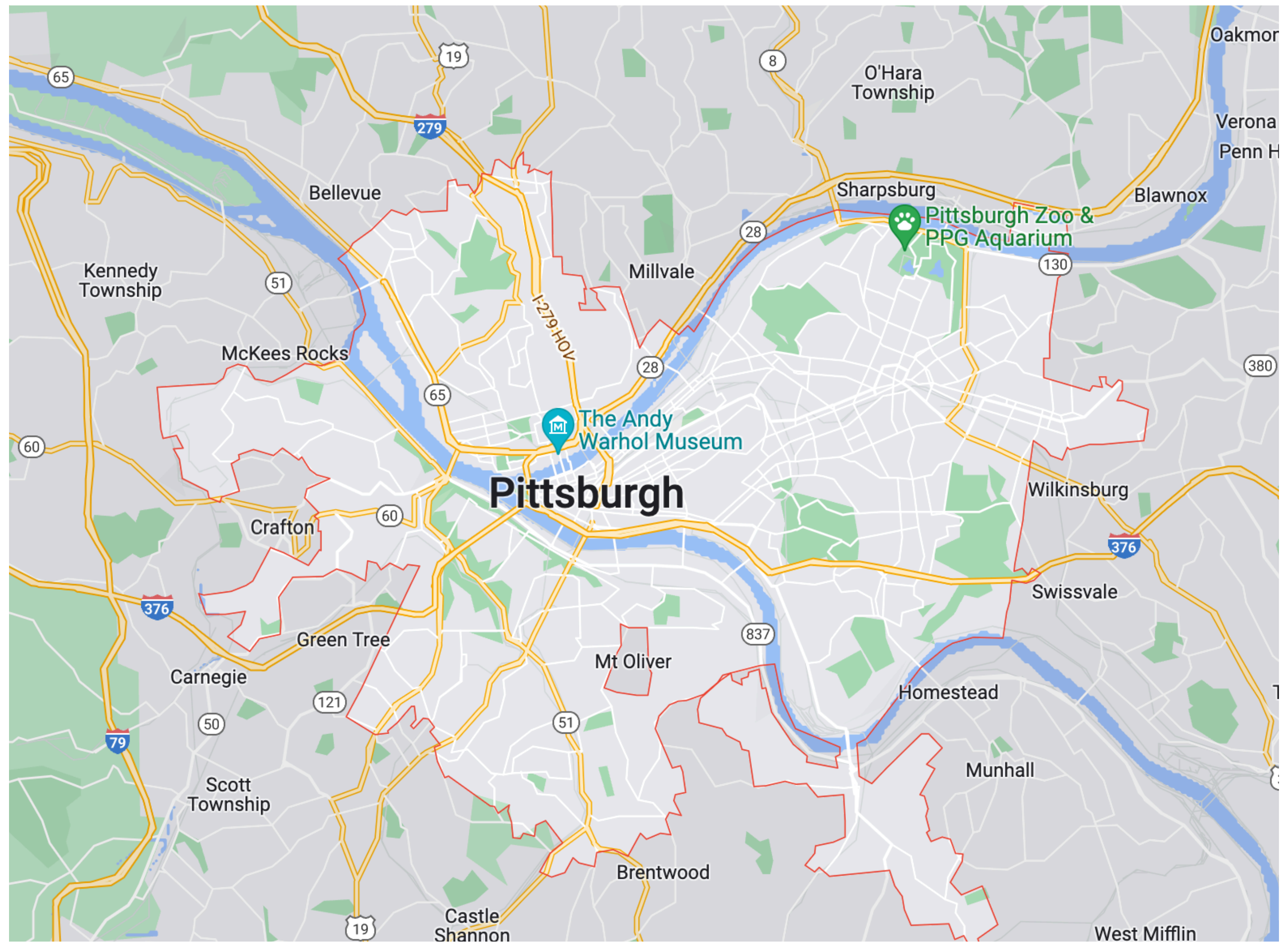
# Pittsburgh vs. Seoul

# Pittsburgh vs. Seoul

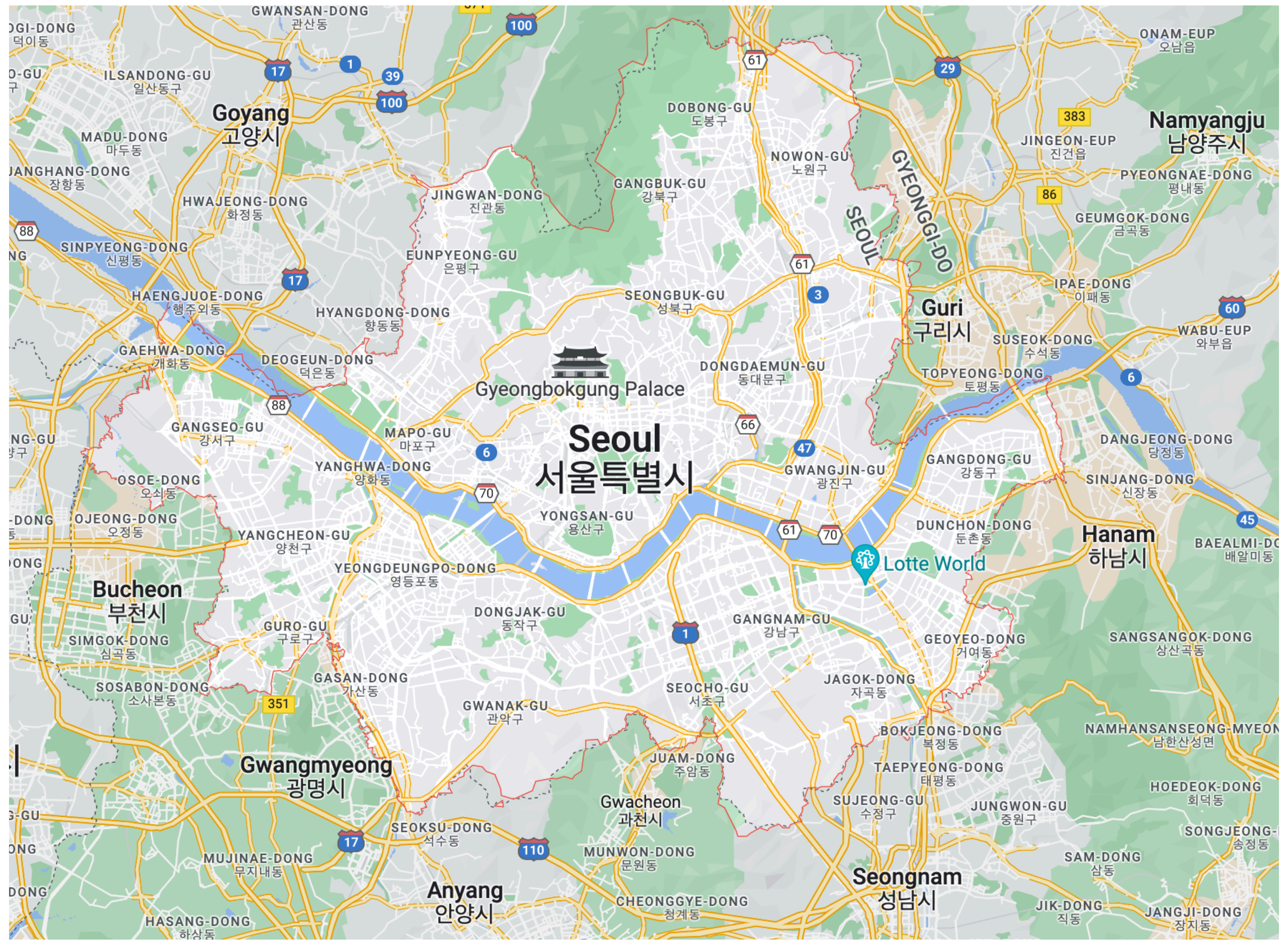
- ***Geography?***

# Pittsburgh vs. Seoul

- ***Geography?***
- Pittsburgh vs. Seoul
  - Shape
  - Structure



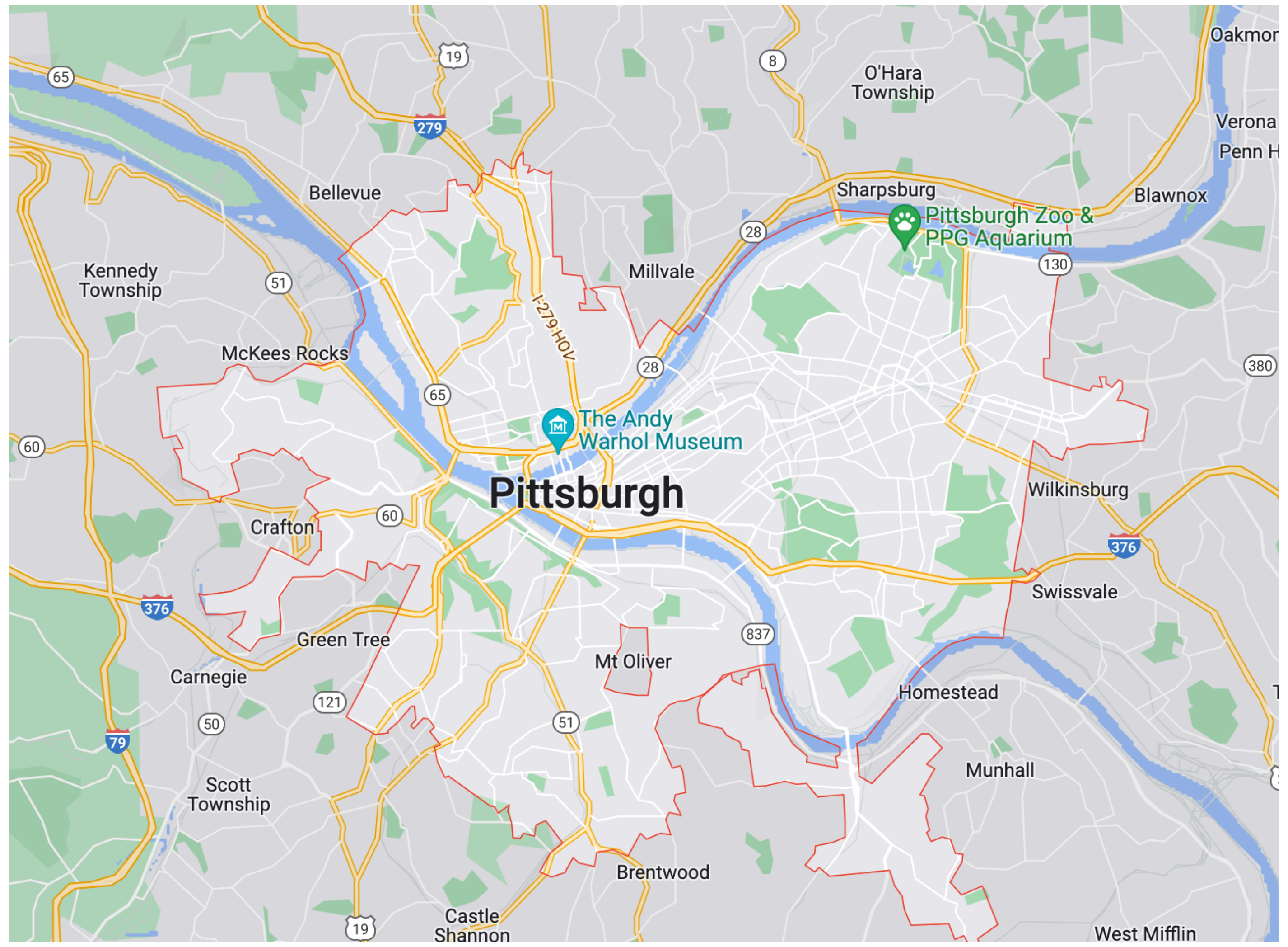

Pittsburgh, Pennsylvania, USA (Credit: Google Maps)



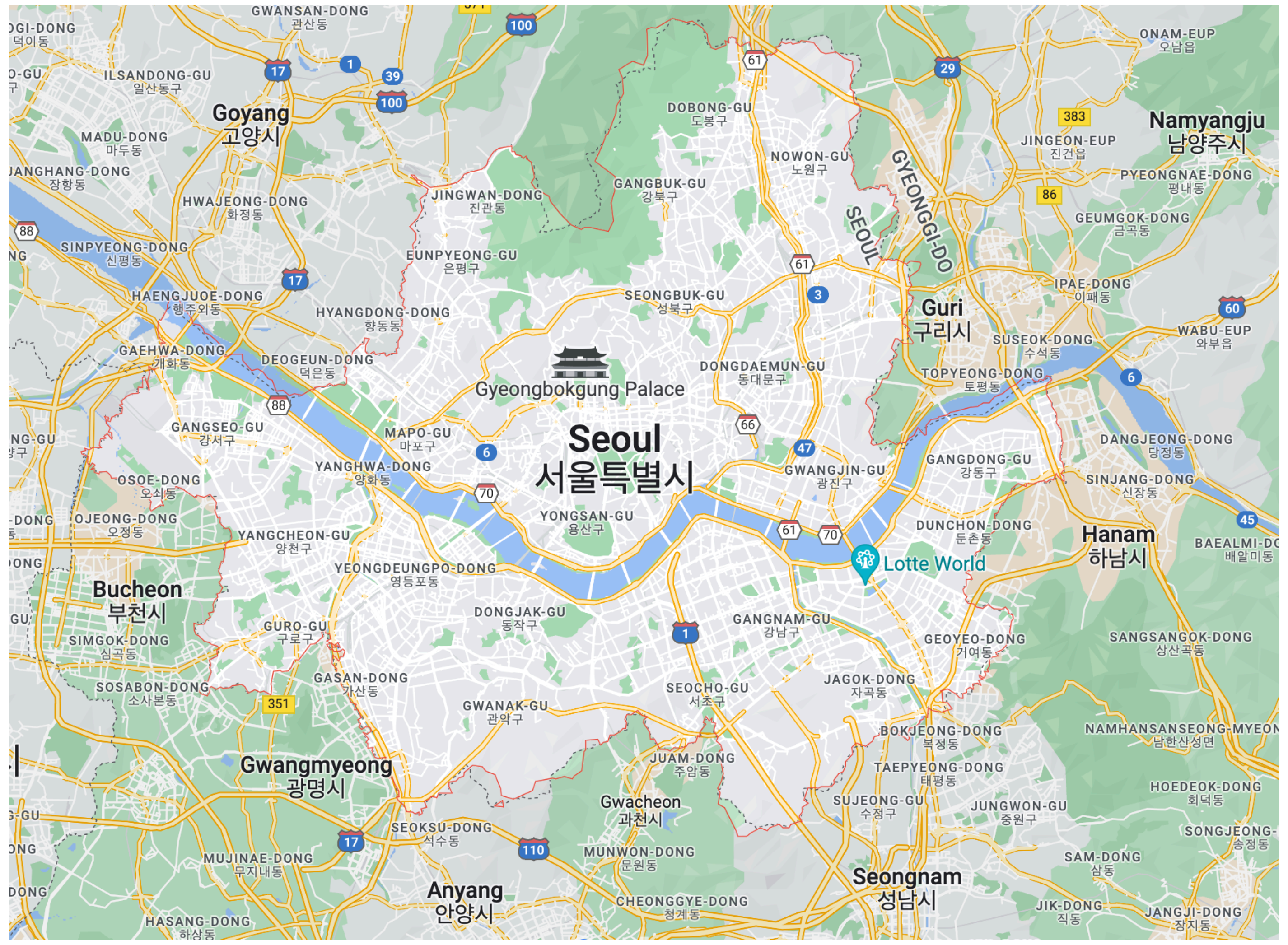

Seoul, South Korea

# Pittsburgh vs. Seoul

- ***Geography?***
- Pittsburgh vs. Seoul
  - Shape
  - Structure
- **Classification**




Pittsburgh, Pennsylvania, USA (Credit: Google Maps)



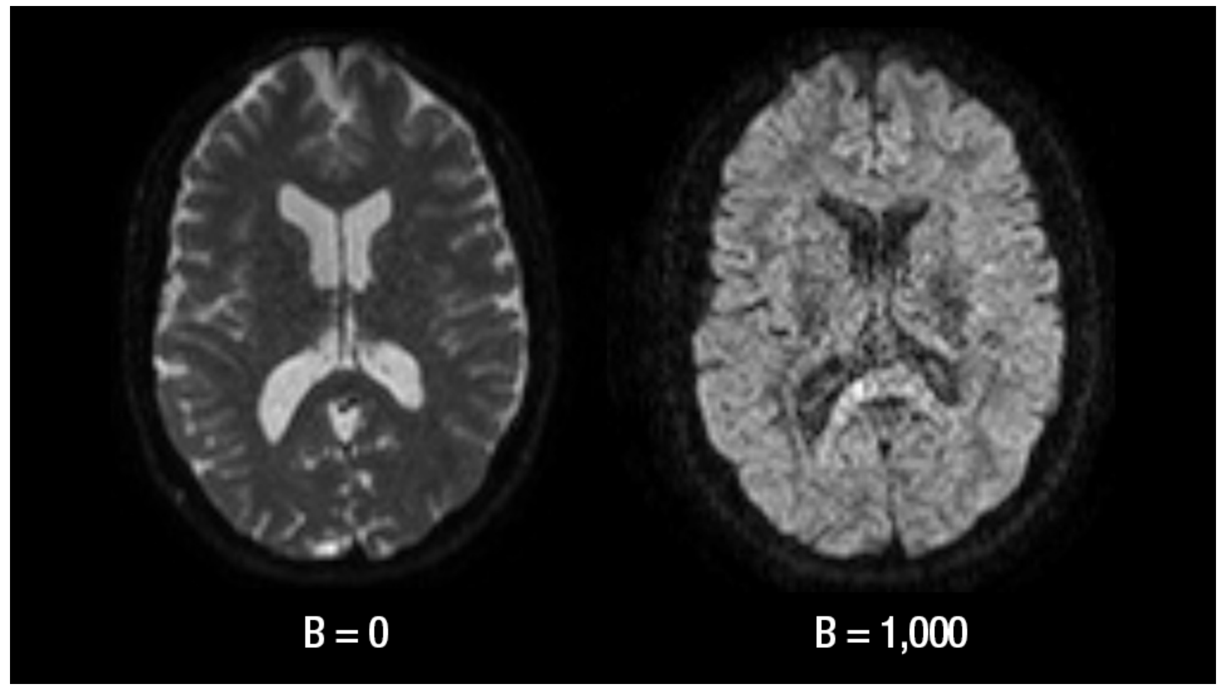

Seoul, South Korea

# Diffusion Tensor Imaging (DTI)

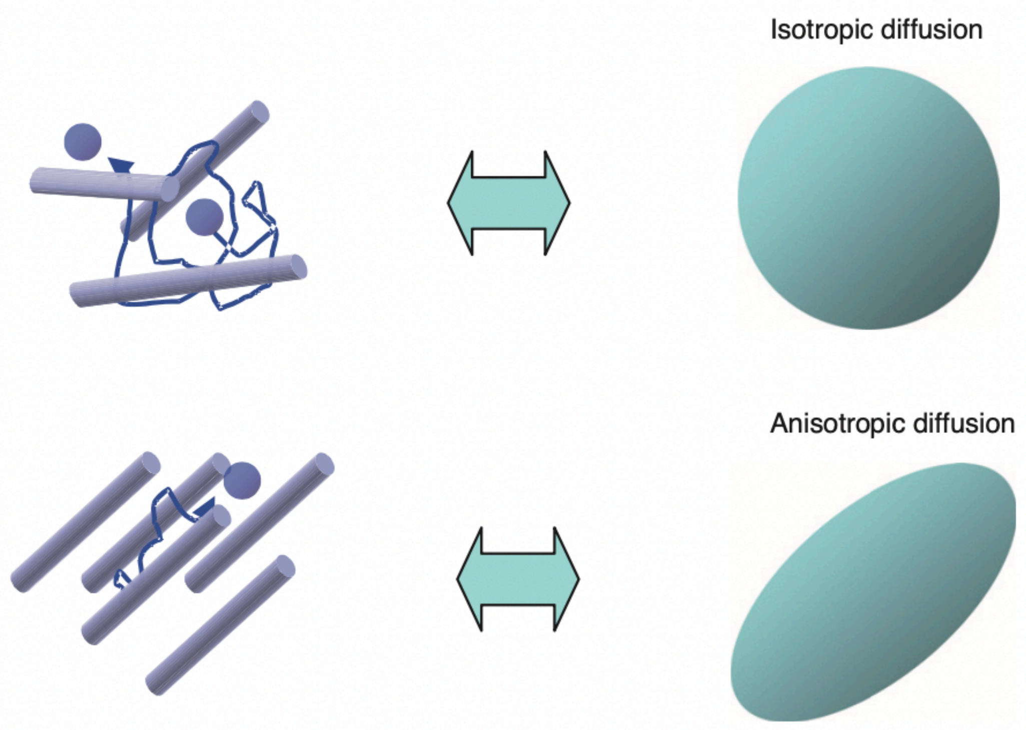
- DTI studies white matter of the brain
- Diffusion-Weighted Imaging (DWI) shows water movement along the tissues and models the difference of signals between baseline image and DWI image at b-value (i.e., b=1000)

$$\ln\left(\frac{S}{S_0}\right) = -\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D = -bD$$

- The diffusion tensor D is used to measure the degree of anisotropy and structural orientation that characterizes DTI




Credit: imagilys.com




Credit. Mori 2007

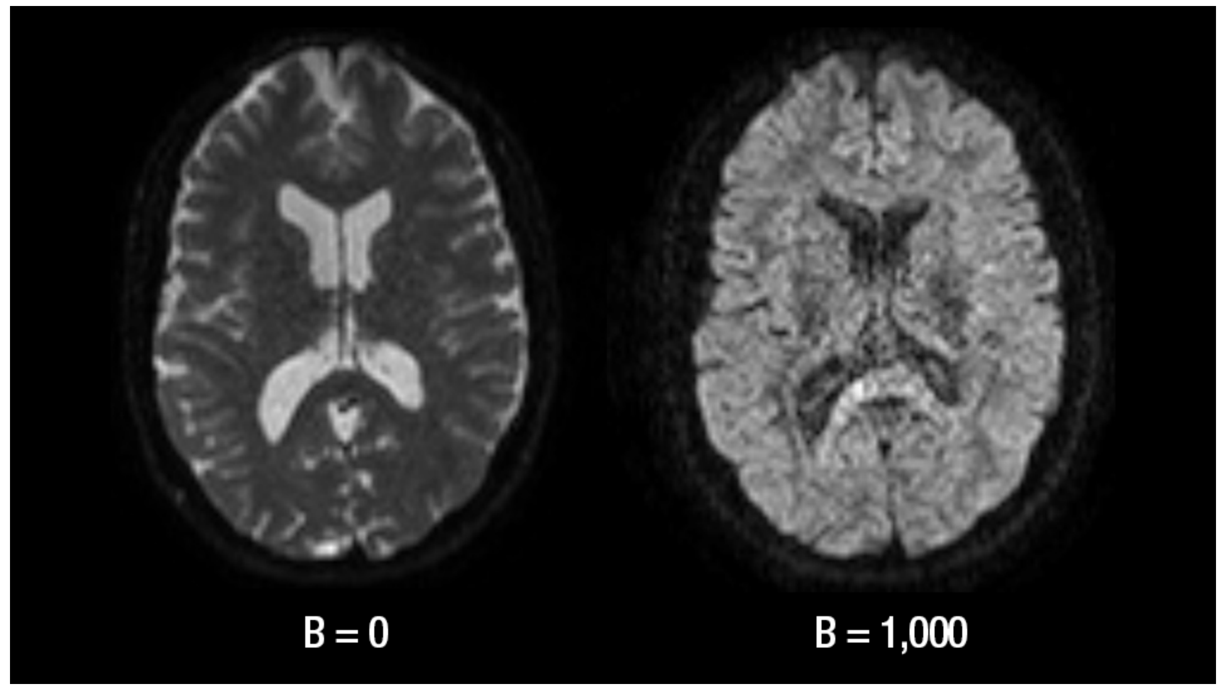
# Diffusion Tensor Imaging (DTI)

- DTI studies white matter of the brain

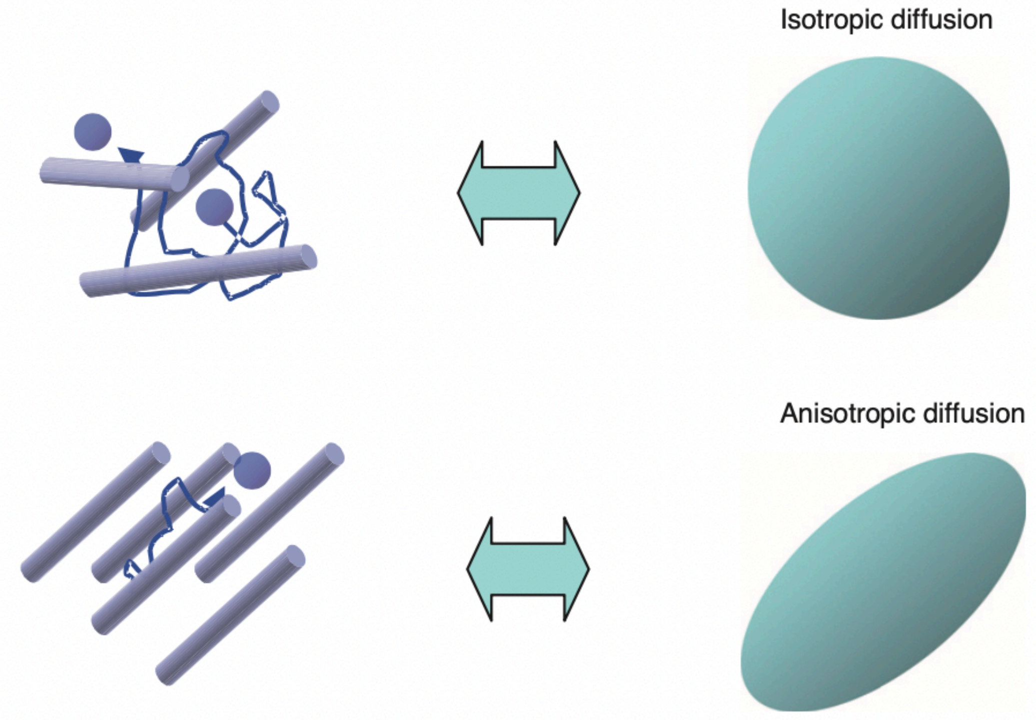
- Diffusion-Weighted Imaging (DWI) shows water movement along the tissues and models the difference of signals between baseline image and DWI image at b-value (i.e., b=1000)

$$\ln\left(\frac{S}{S_0}\right) = -\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D = -bD$$

- The diffusion tensor D is used to measure the degree of anisotropy and structural orientation that characterizes DTI




Credit: imagilys.com



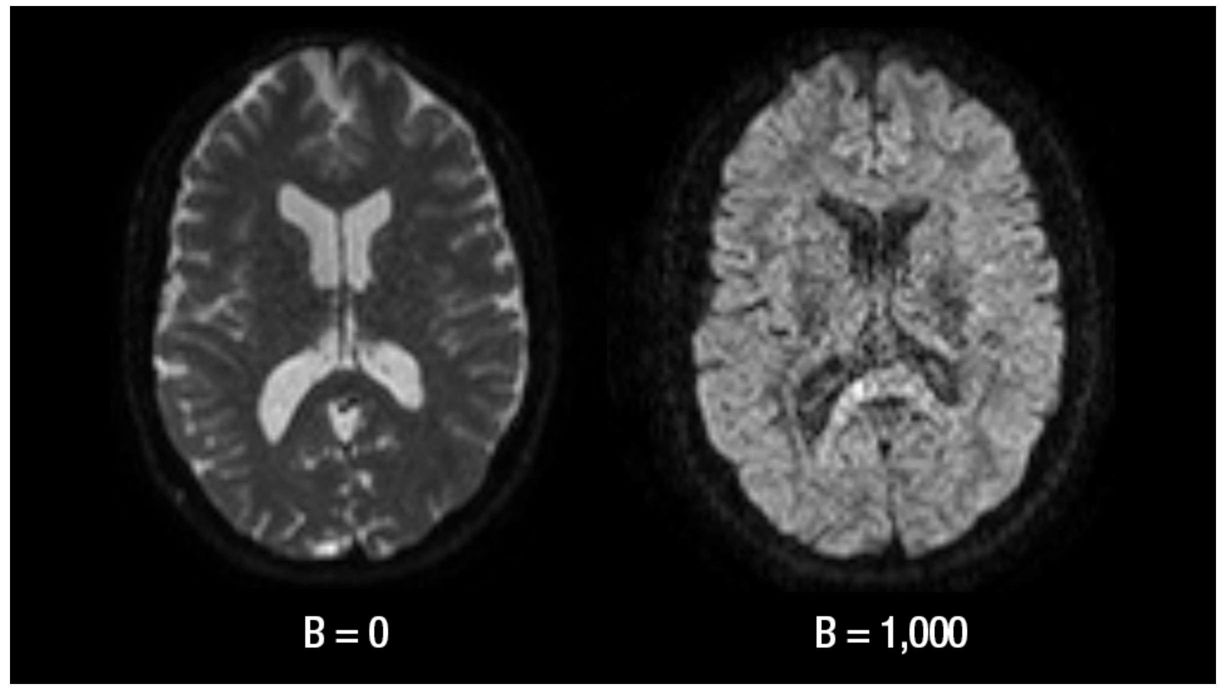

Credit. Mori 2007

# Diffusion Tensor Imaging (DTI)

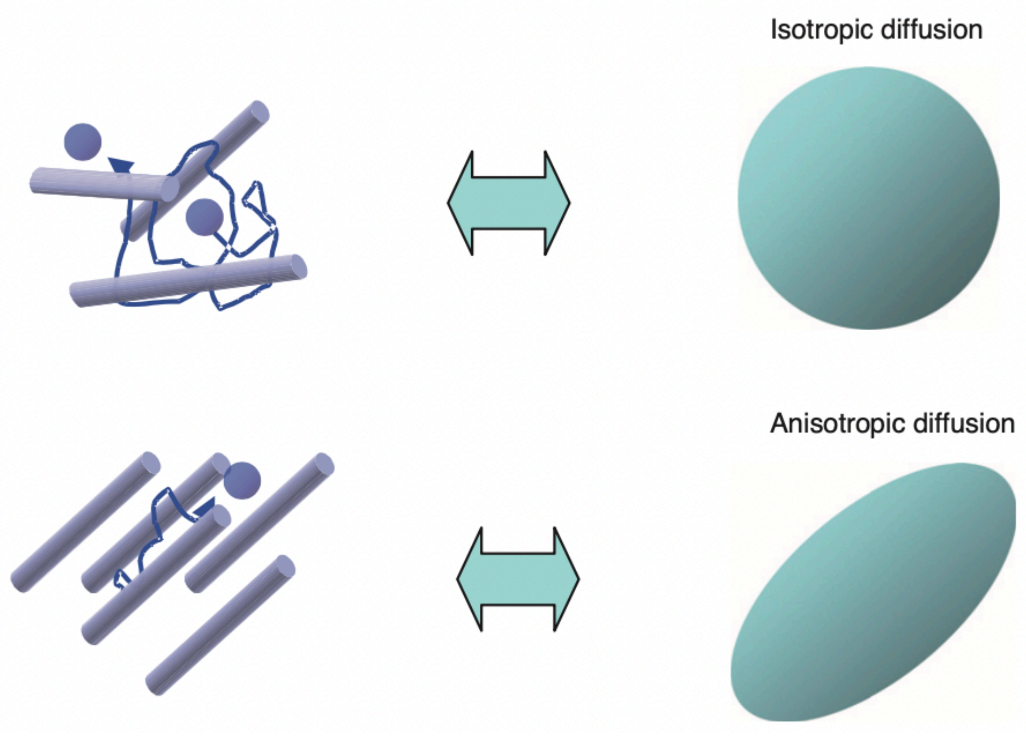
- DTI studies white matter of the brain
- Diffusion-Weighted Imaging (DWI) shows water movement along the tissues and models the difference of signals between baseline image and DWI image at b-value (i.e., b=1000)

$$\ln\left(\frac{S}{S_0}\right) = -\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D = -bD$$

- The diffusion tensor D is used to measure the degree of anisotropy and structural orientation that characterizes DTI




Credit: imagilys.com



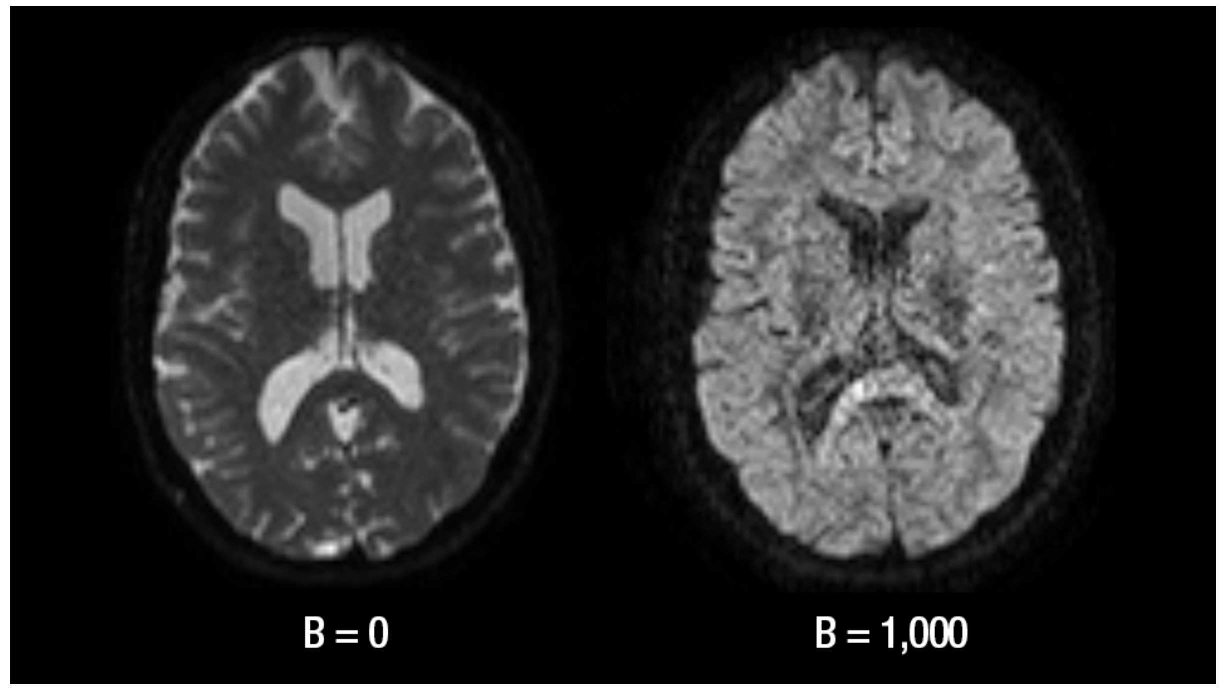

Credit. Mori 2007

# Diffusion Tensor Imaging (DTI)

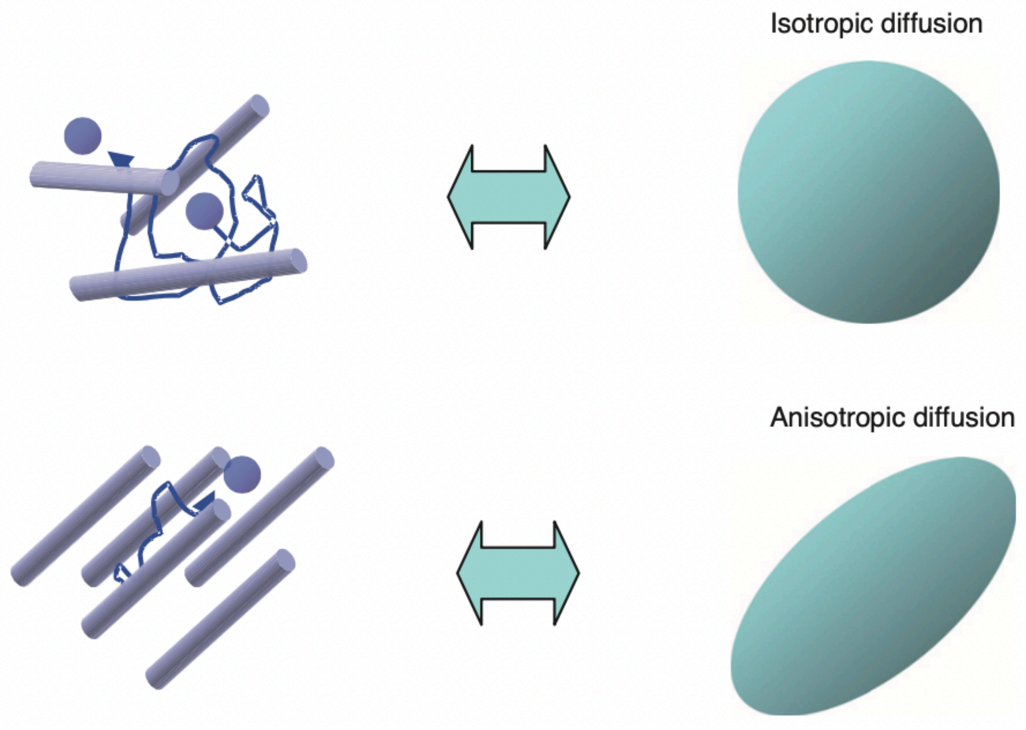
- DTI studies white matter of the brain
- Diffusion-Weighted Imaging (DWI) shows water movement along the tissues and models the difference of signals between baseline image and DWI image at b-value (i.e., b=1000)

$$\ln\left(\frac{S}{S_0}\right) = -\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D = -bD$$

- The diffusion tensor D is used to measure the degree of anisotropy and structural orientation that characterizes DTI




Credit: imagilys.com



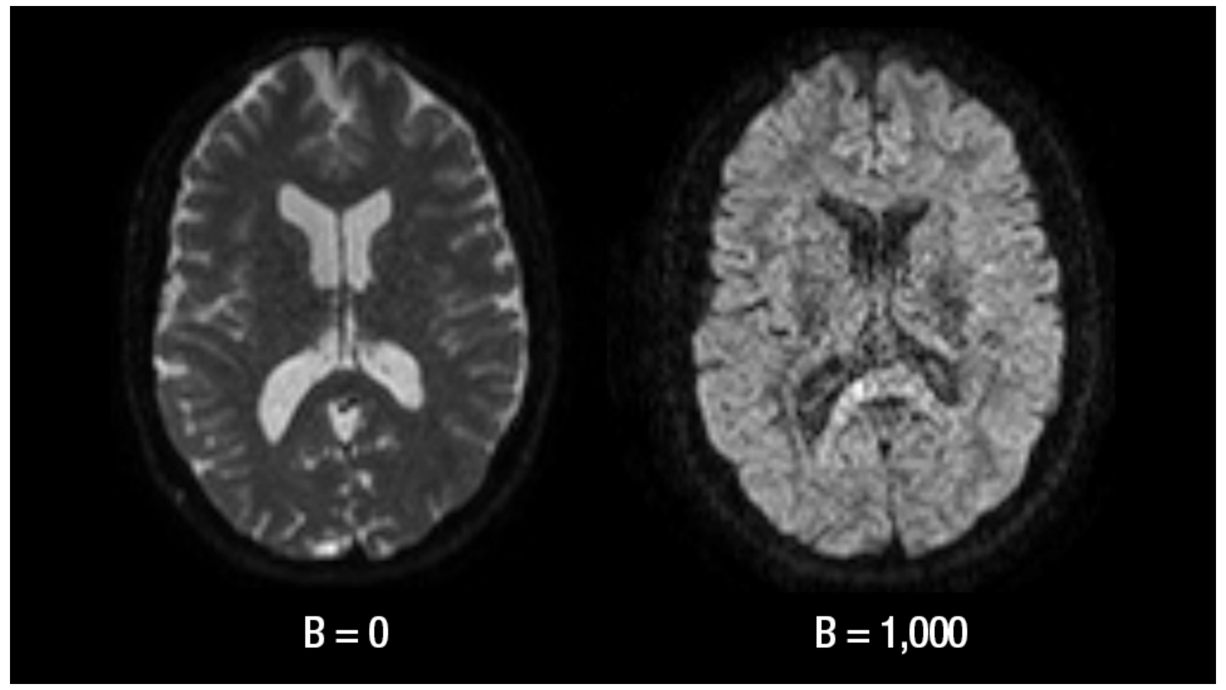

Credit. Mori 2007

# Diffusion Tensor Imaging (DTI)

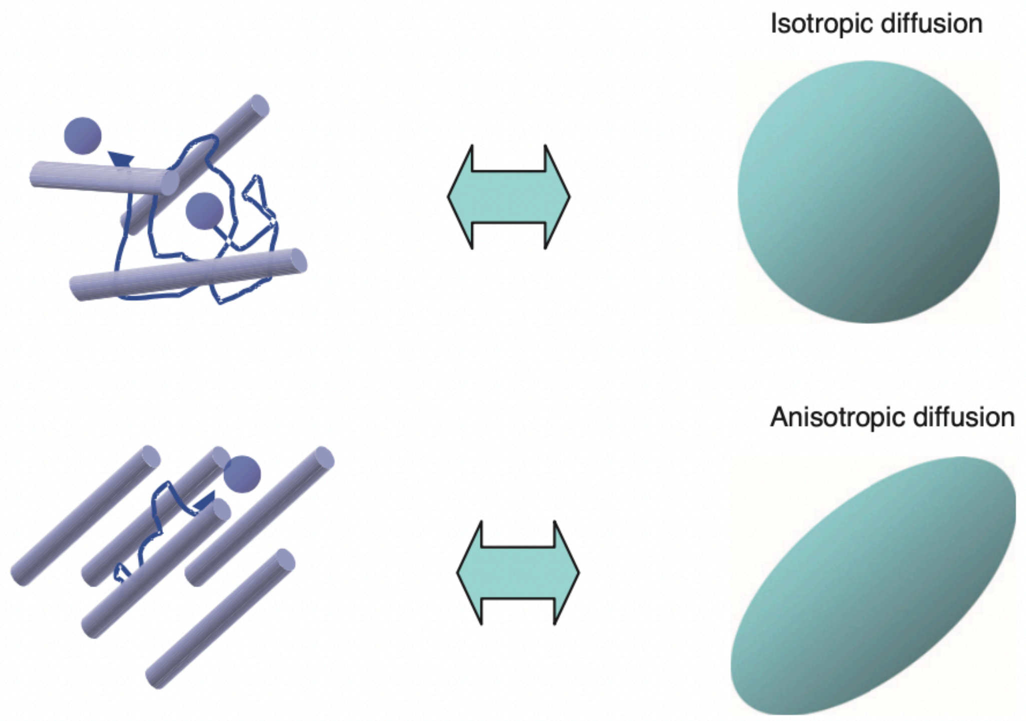
- DTI studies white matter of the brain
- Diffusion-Weighted Imaging (DWI) shows water movement along the tissues and models the difference of signals between baseline image and DWI image at b-value (i.e., b=1000)

$$\ln\left(\frac{S}{S_0}\right) = -\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D = -bD$$

- The diffusion tensor D is used to measure the degree of anisotropy and structural orientation that characterizes DTI



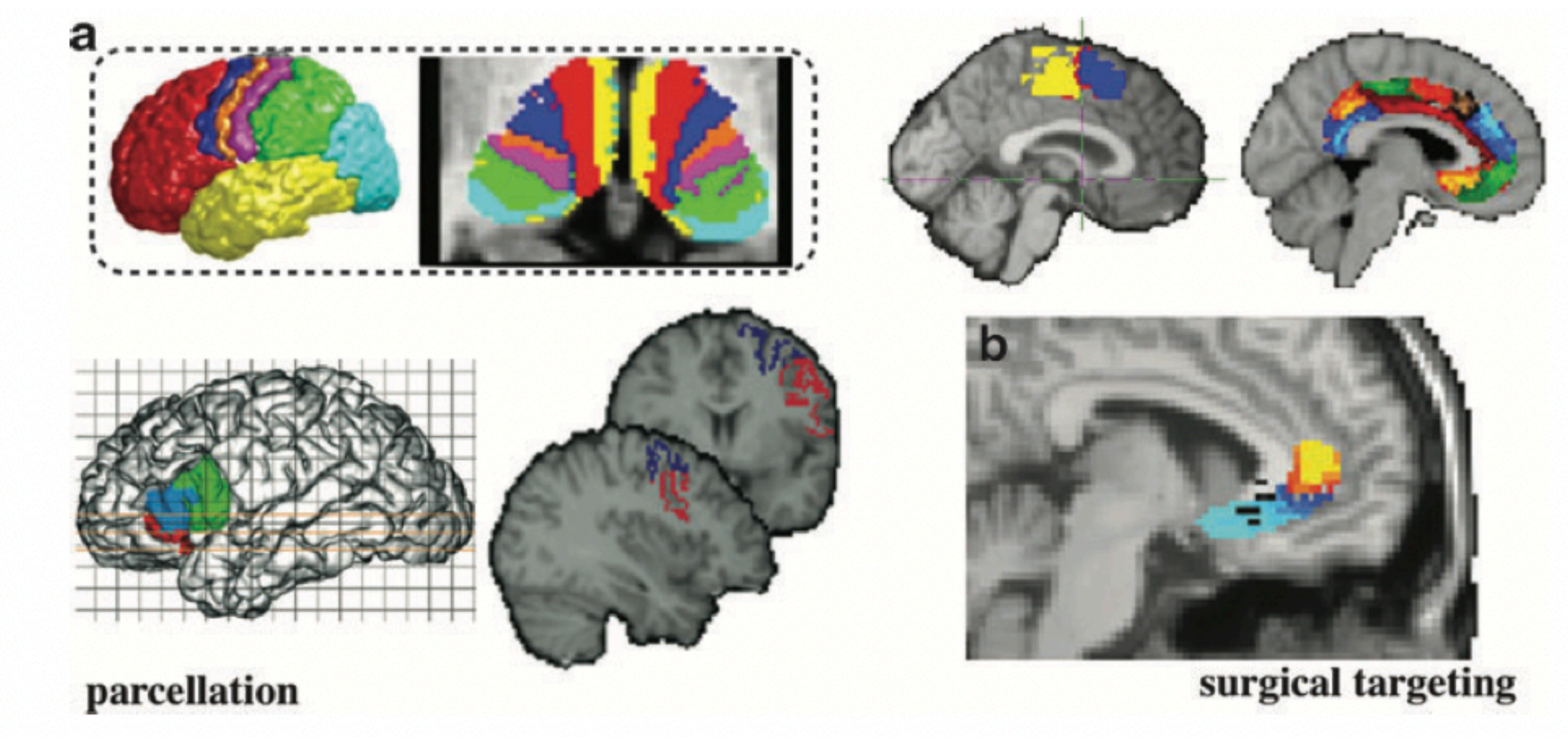

Credit: imagilys.com




Credit. Mori 2007

# Existing DTI studies

- Both normal subjects and patients with brain tumors, Alzheimer's, schizophrenia, etc.
- ROI Analysis, Voxel-Based Analysis, Tract-Based Spatial Statistics (e.g., tractography)
- Manual annotation of region by experts




Credit: Jbabdi and Johansen-Berg 2011

# New question

# New question

- An important issue was not addressed in DTI studies.

# New question

- An important issue was not addressed in DTI studies.

**Can we define a measure that distinguishes different groups?**

# New question

- An important issue was not addressed in DTI studies.

**Can we define a measure that distinguishes different groups?**

- Assumption: If we find a measure that defines some good feature in DTI, then we don't have to depend on previous regions of interest (ROI) that requires manual feature selection.

# New question

- An important issue was not addressed in DTI studies.

**Can we define a measure that distinguishes different groups?**

- Assumption: If we find a measure that defines some good feature in DTI, then we don't have to depend on previous regions of interest (ROI) that requires manual feature selection.

➢ **Holistic approach to brain images searching for patterns**

# Contents

# Mathematics of DTI

- Diffusion process by an ellipsoid (Basser et al., 1994)
- Diffusion Tensor — A symmetric 3x3 matrix for each voxel
- Shape and orientation — eigenvalues and corresponding eigenvectors

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]$$

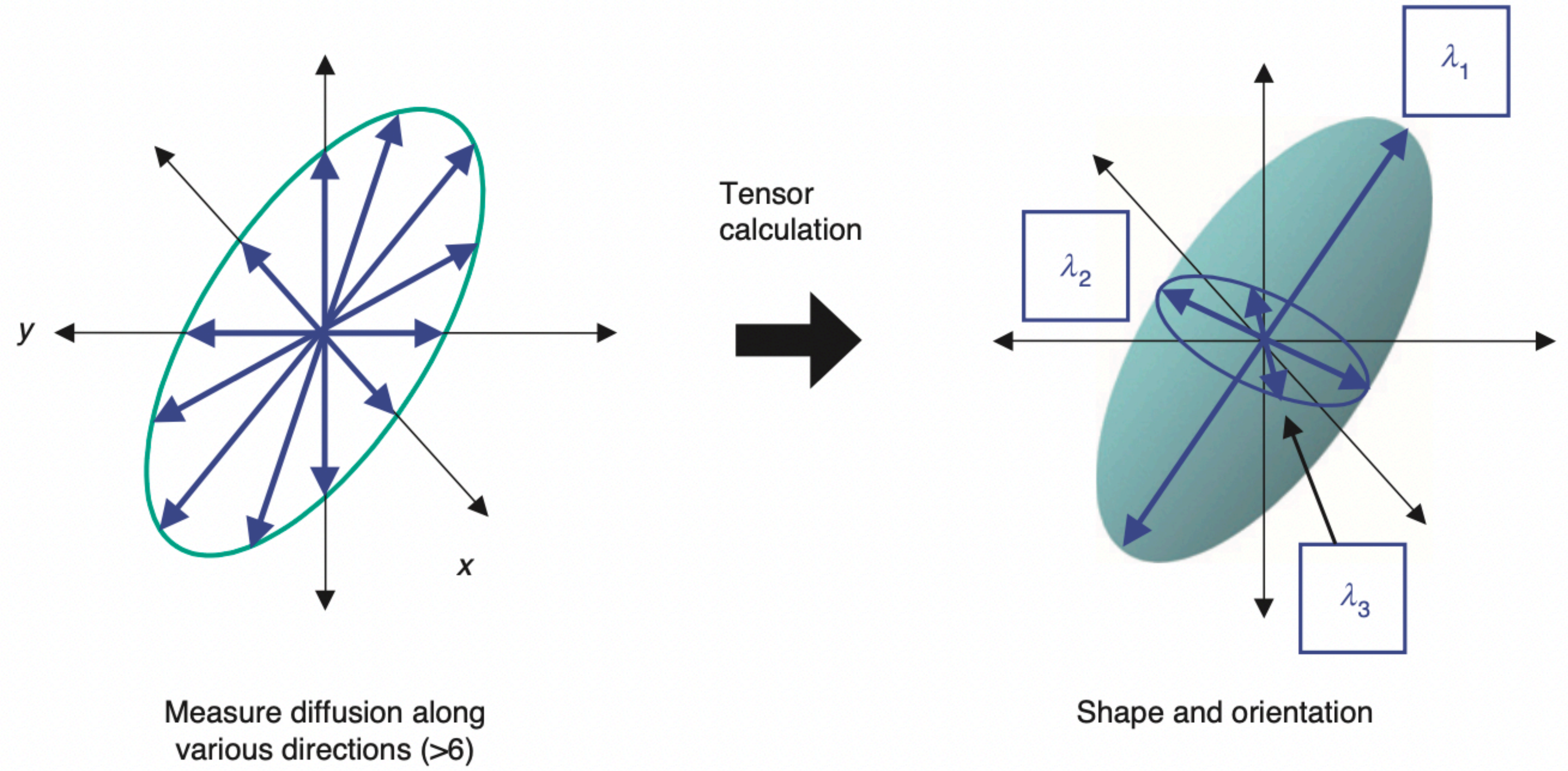



Fig. 5.2 A diffusion ellipsoid can be fully characterized from diffusion measurements along six independent axes.

Credit: Mori 2007

# Mathematics of DTI

$$FA = \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$$

where $\bar{\lambda} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} = MD$

- Diffusion process by an ellipsoid (Basser et al., 1994)
- Diffusion Tensor — A symmetric 3x3 matrix for each voxel
- Shape and orientation — eigenvalues and corresponding eigenvectors

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]$$

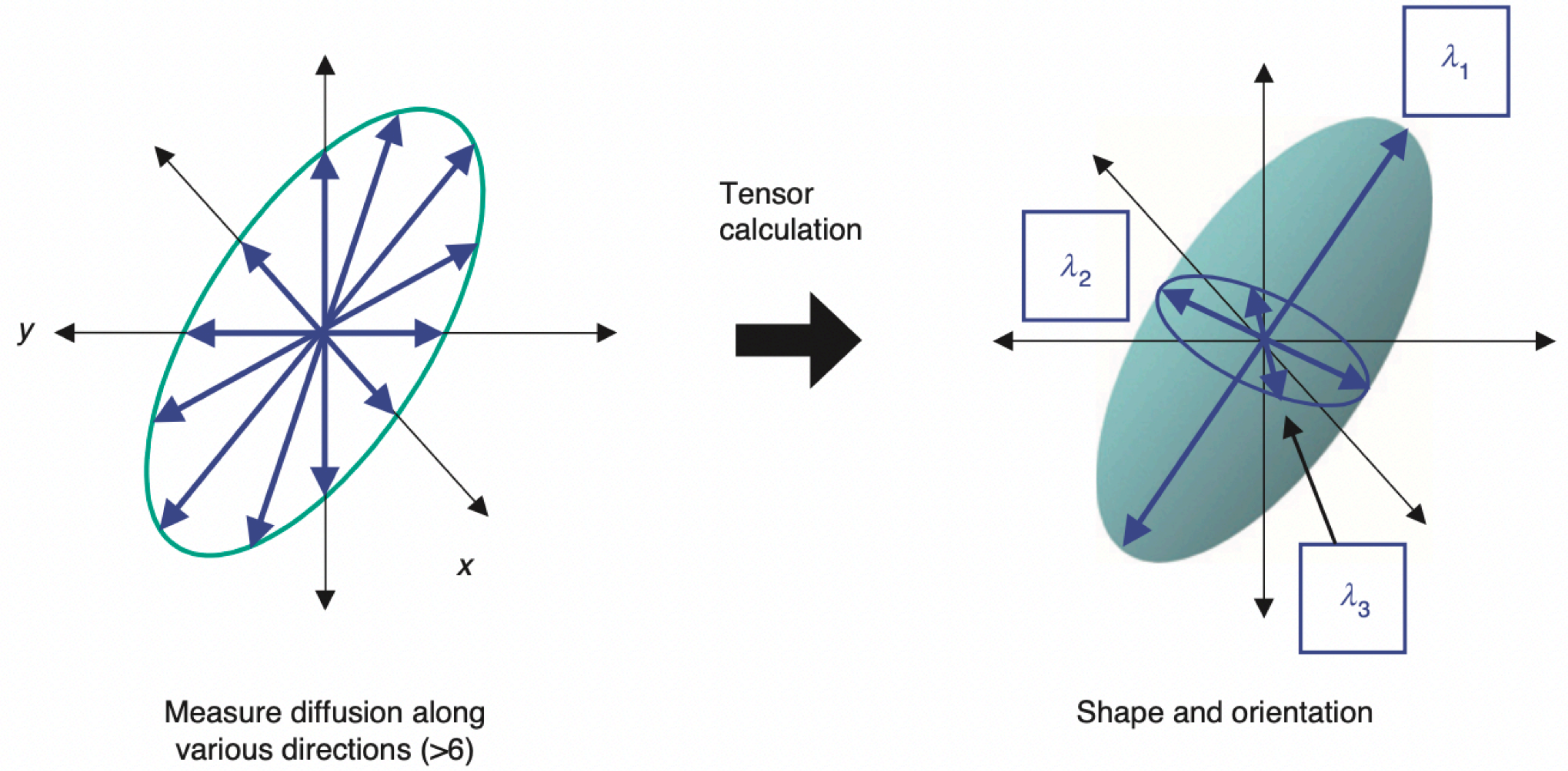



Fig. 5.2 A diffusion ellipsoid can be fully characterized from diffusion measurements along six independent axes.

Credit: Mori 2007

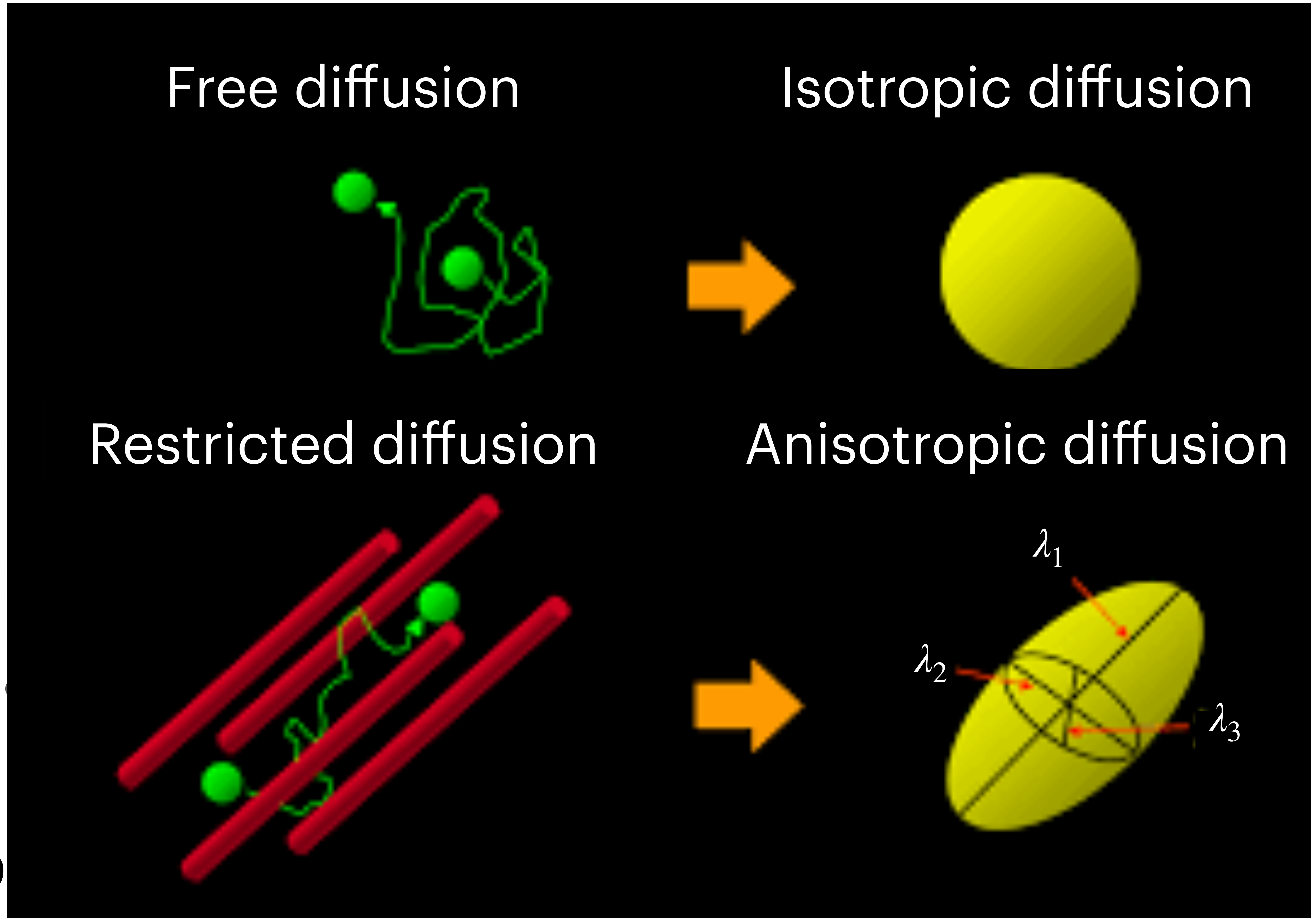
# Mathematics of DTI

$$FA = \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$$

where $\bar{\lambda} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} = MD$

- Diffusion process by an ellipsoid (Basser et al., 1994)
- Diffusion Tensor — A symmetric 3x3 matrix for each voxel
- Shape and orientation — eigenvalues and corresponding eigenvectors

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3]$$

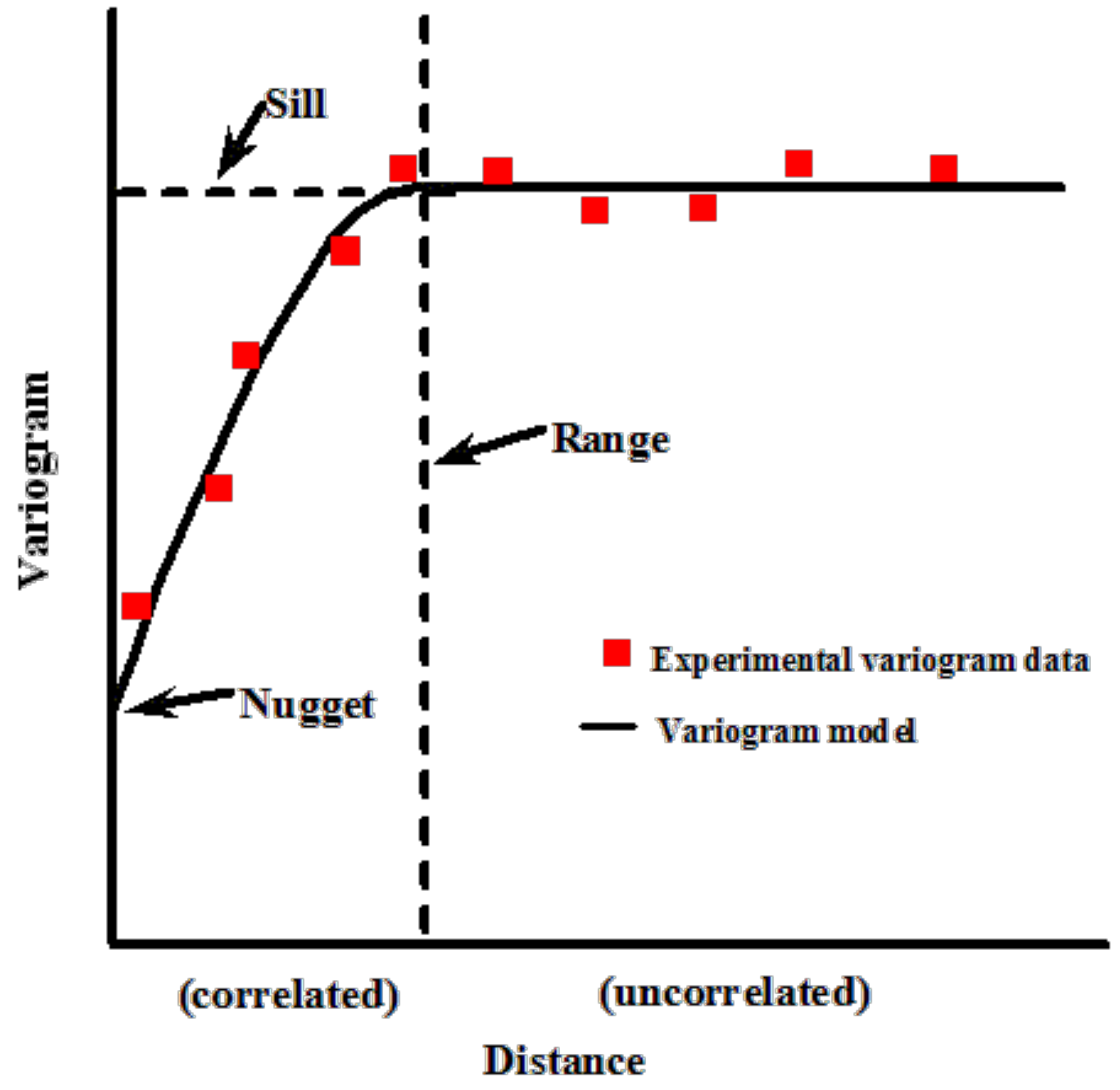


# Ideas for a new measure

# Ideas for a new measure

1. Semi-variogram (Matheron, 1963)

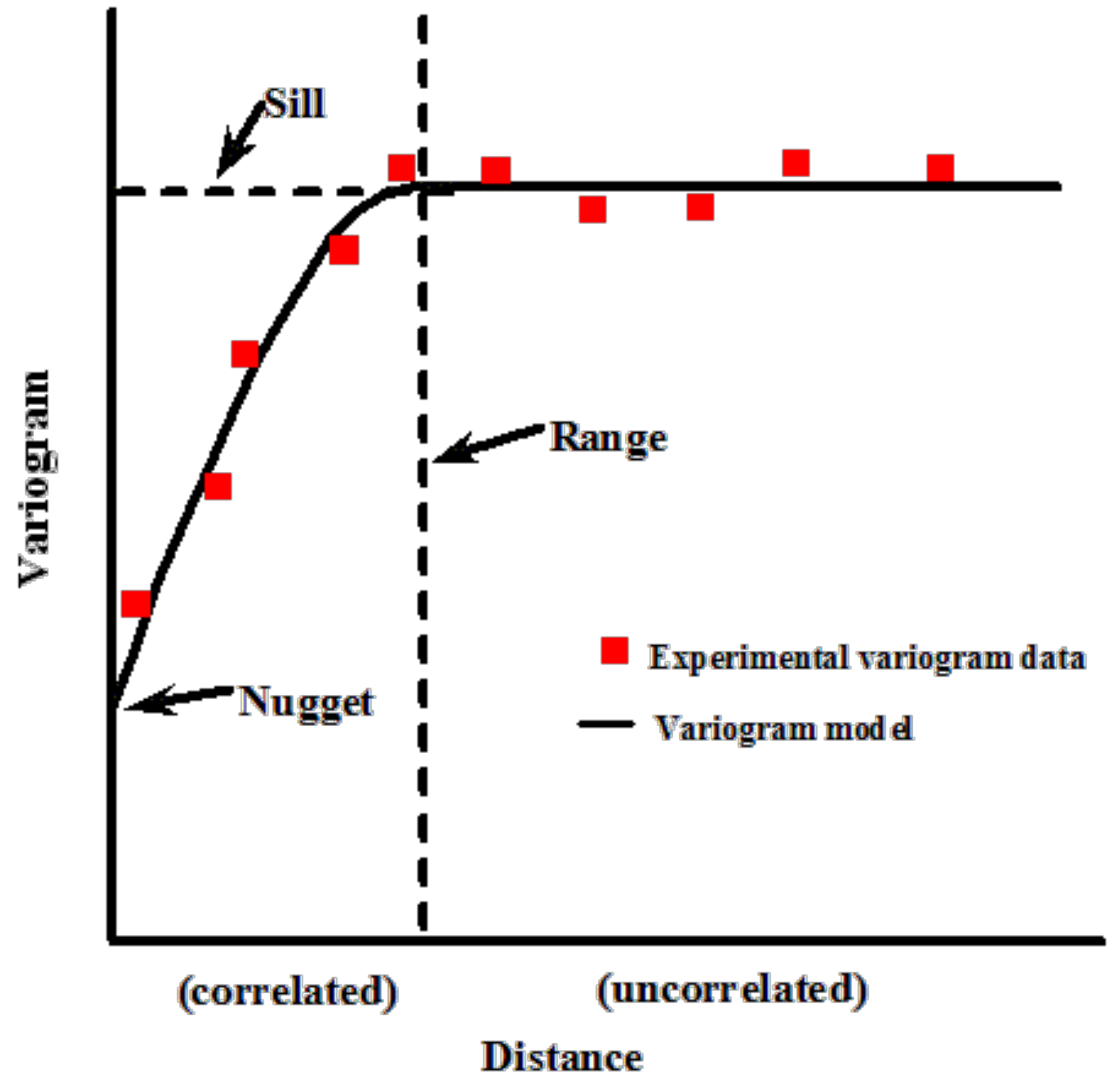
$$\gamma(h) = \frac{1}{2V}\iiint_V [f(M+h) - f(M)]^2 dV$$



# Ideas for a new measure

1. Semi-variogram (Matheron, 1963)

$$\gamma(h) = \frac{1}{2V}\iiint_V [f(M+h) - f(M)]^2 dV$$




2. Polycentric Circle Pooling (Qi et al., 2020)

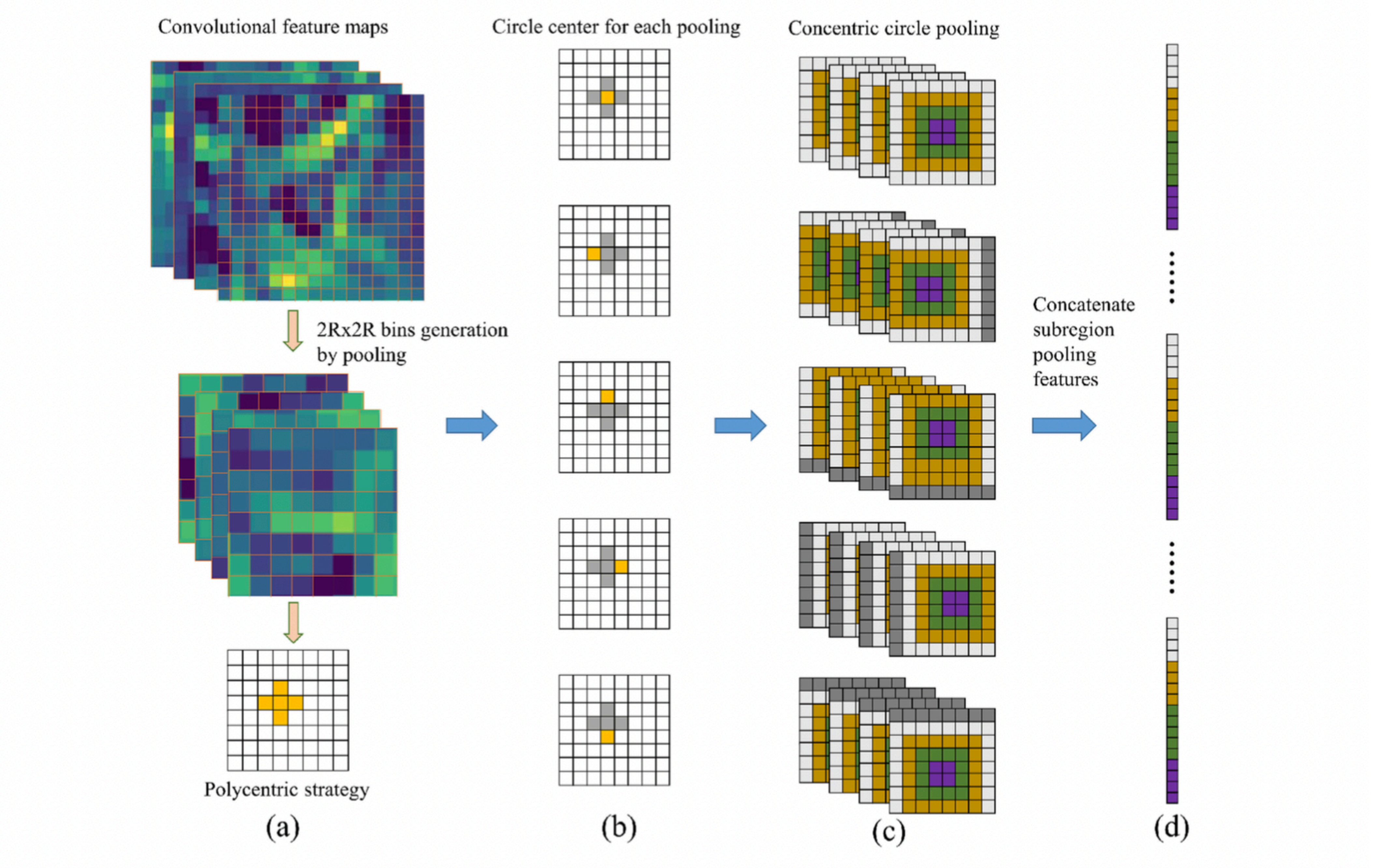
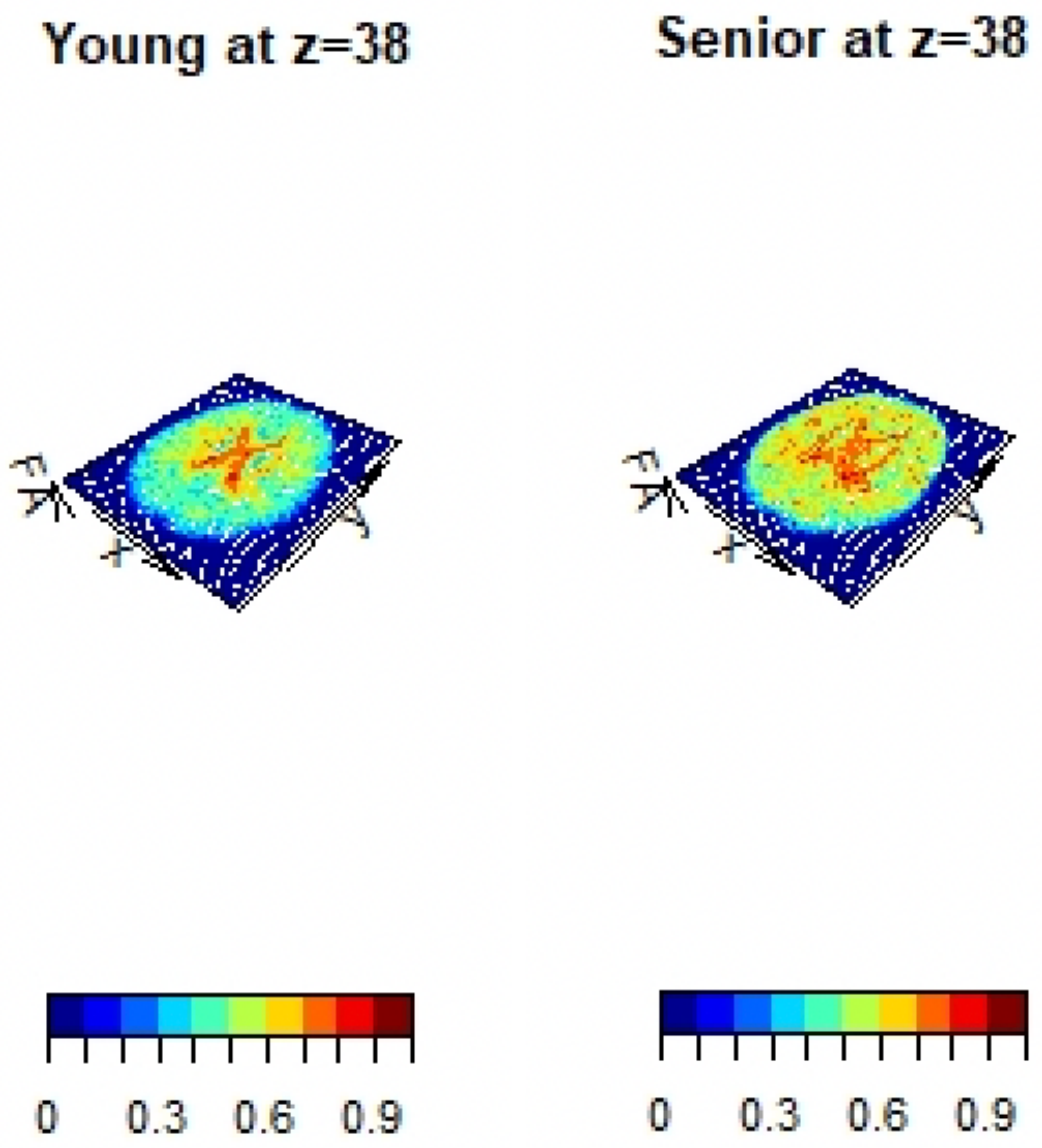
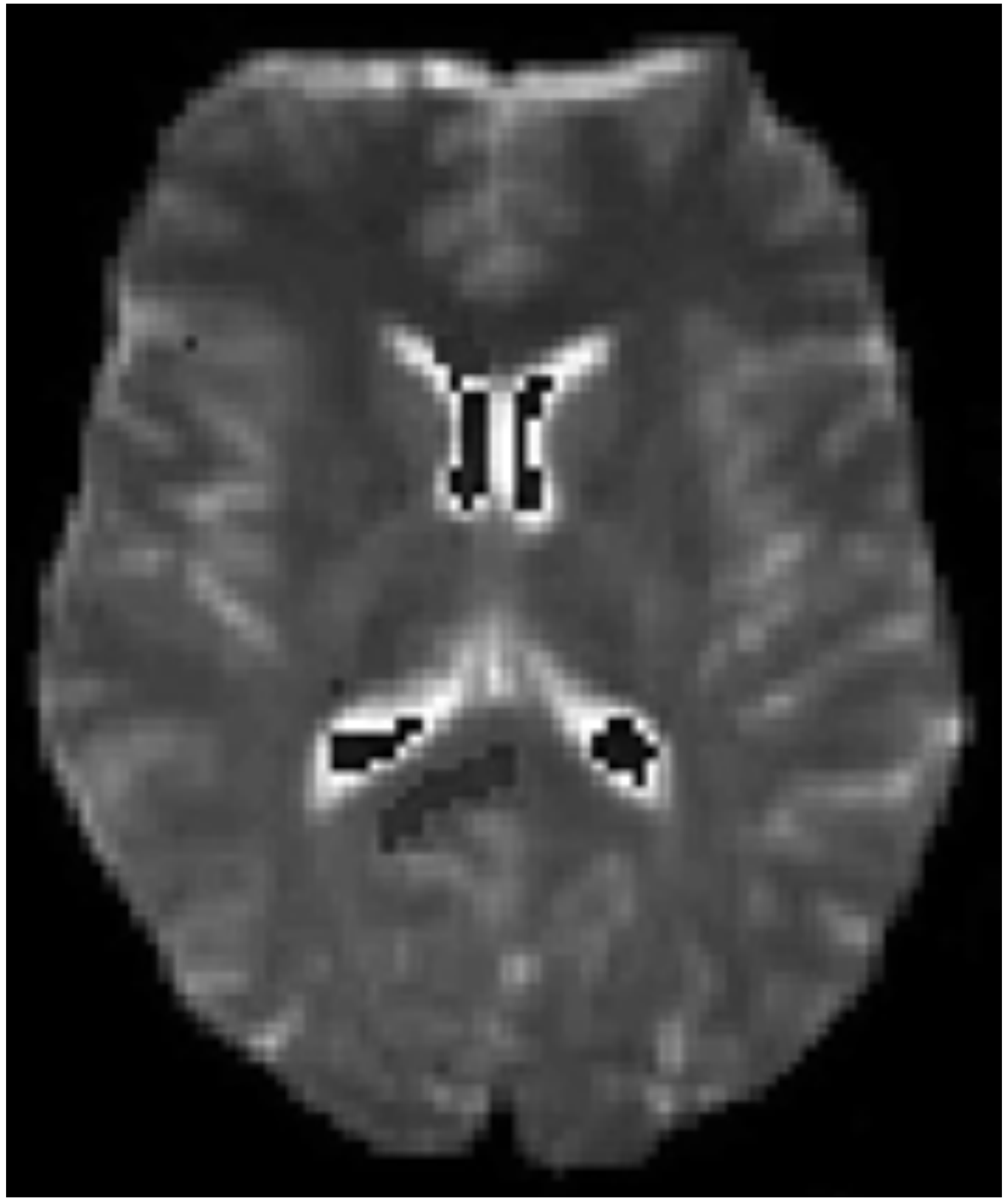



Fig. 3. Example of the pooling strategy in PCP with four-channel maps of size 16 × 16 and five circle centers. The annular subregions are shown in different colors, except for the dark gray indicating the exclusive cells.

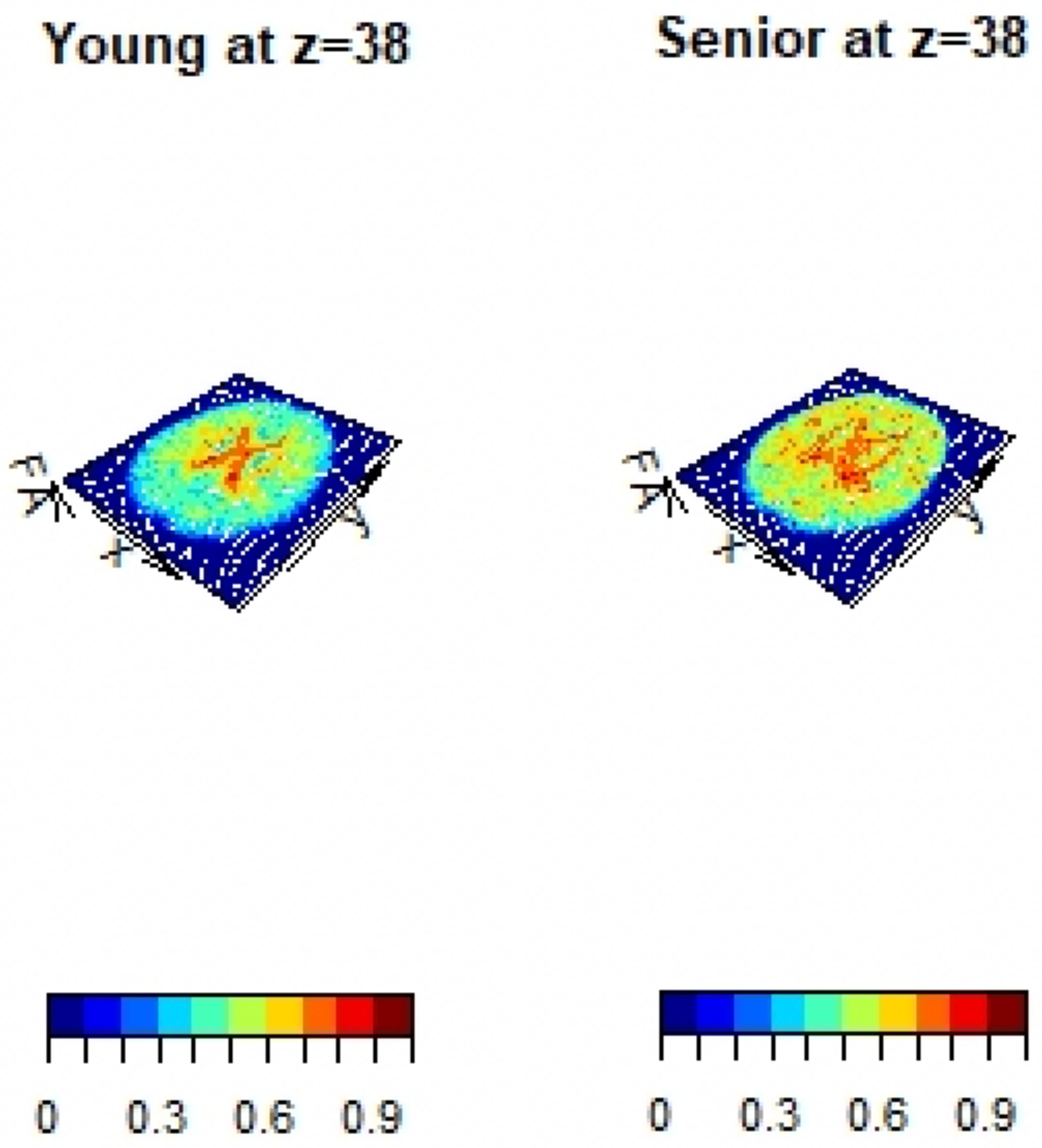
# Central Peripheral Deviation (CPD)



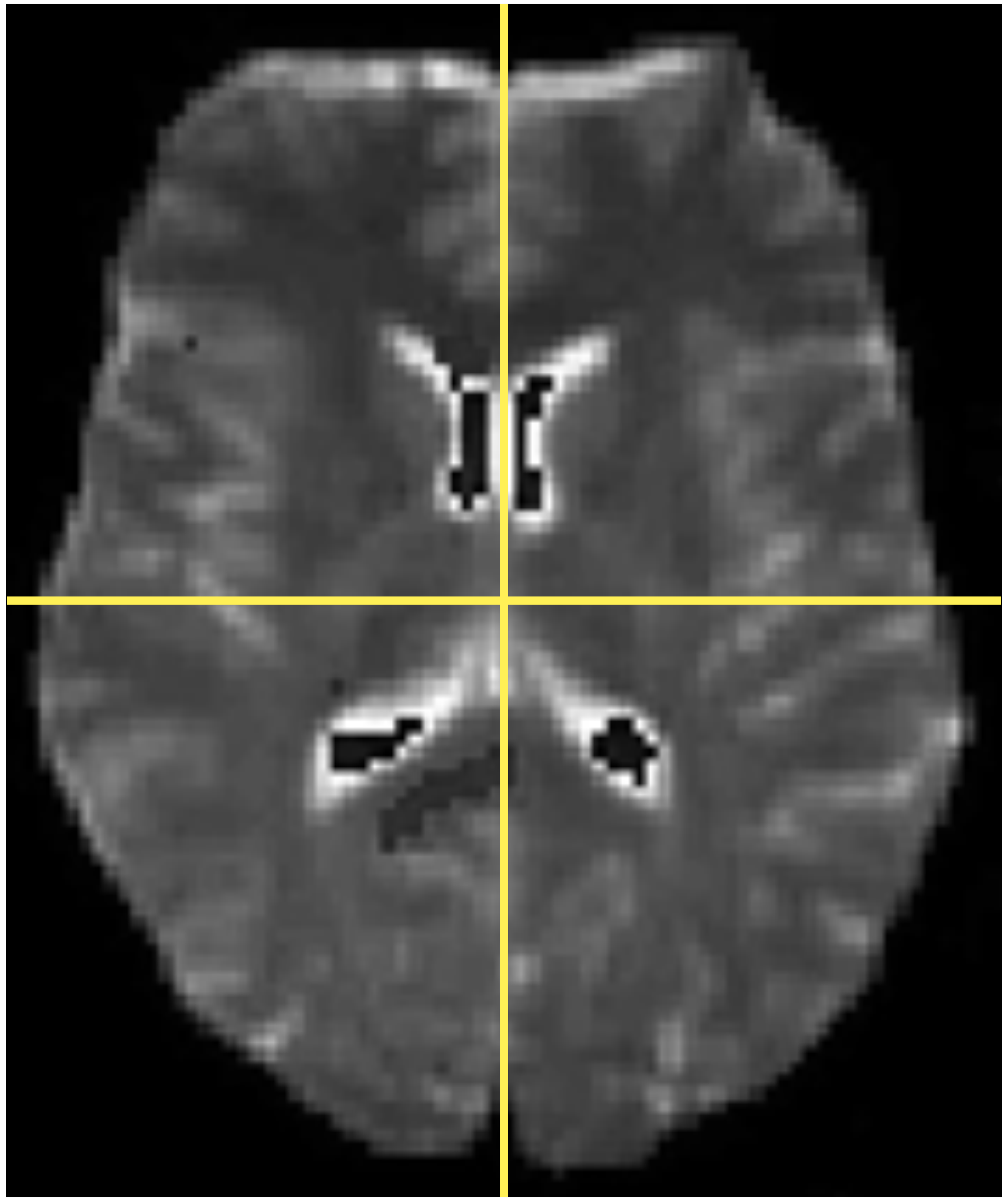

ICBM example: Age group difference in FA



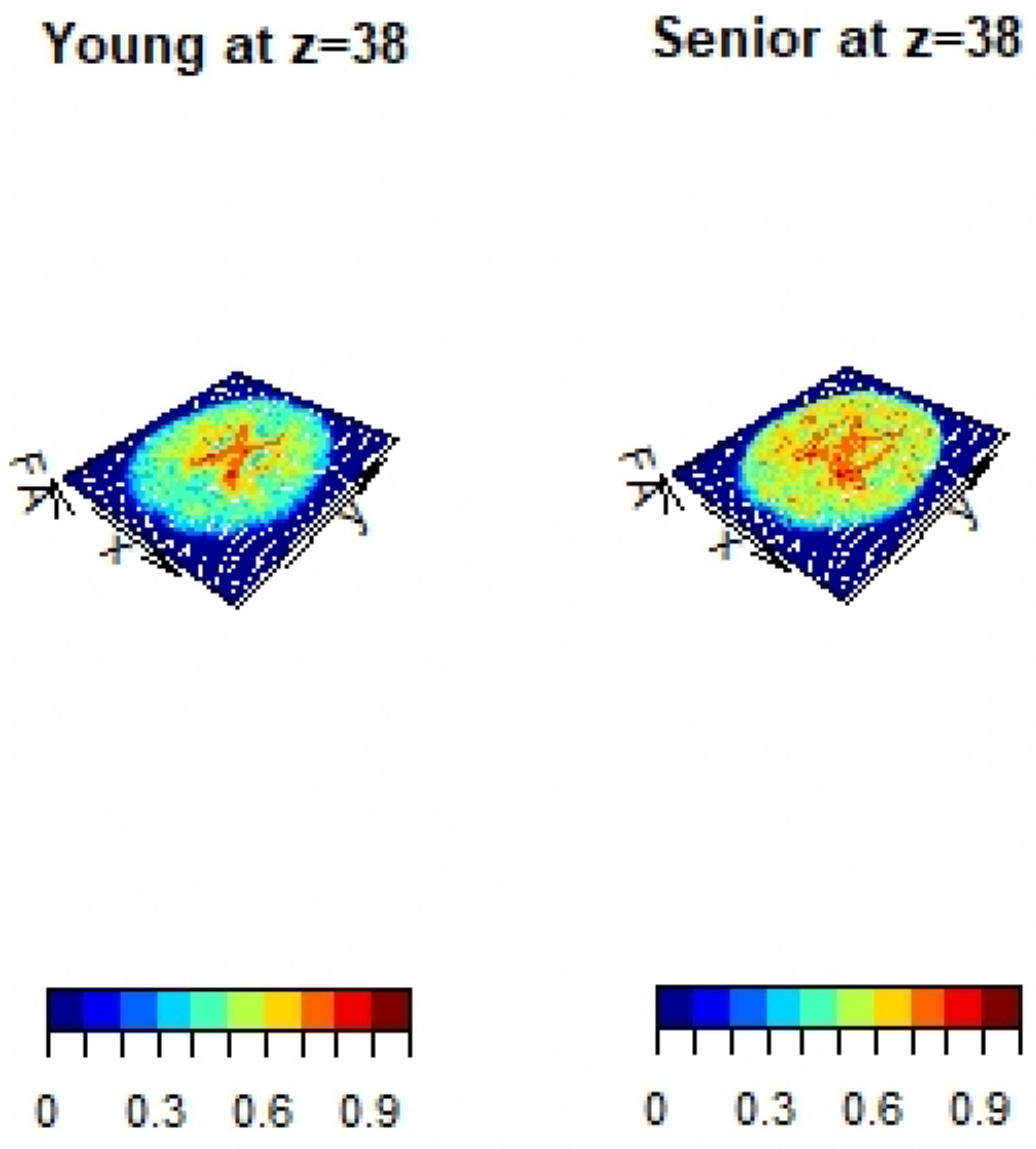
# Central Peripheral Deviation (CPD)



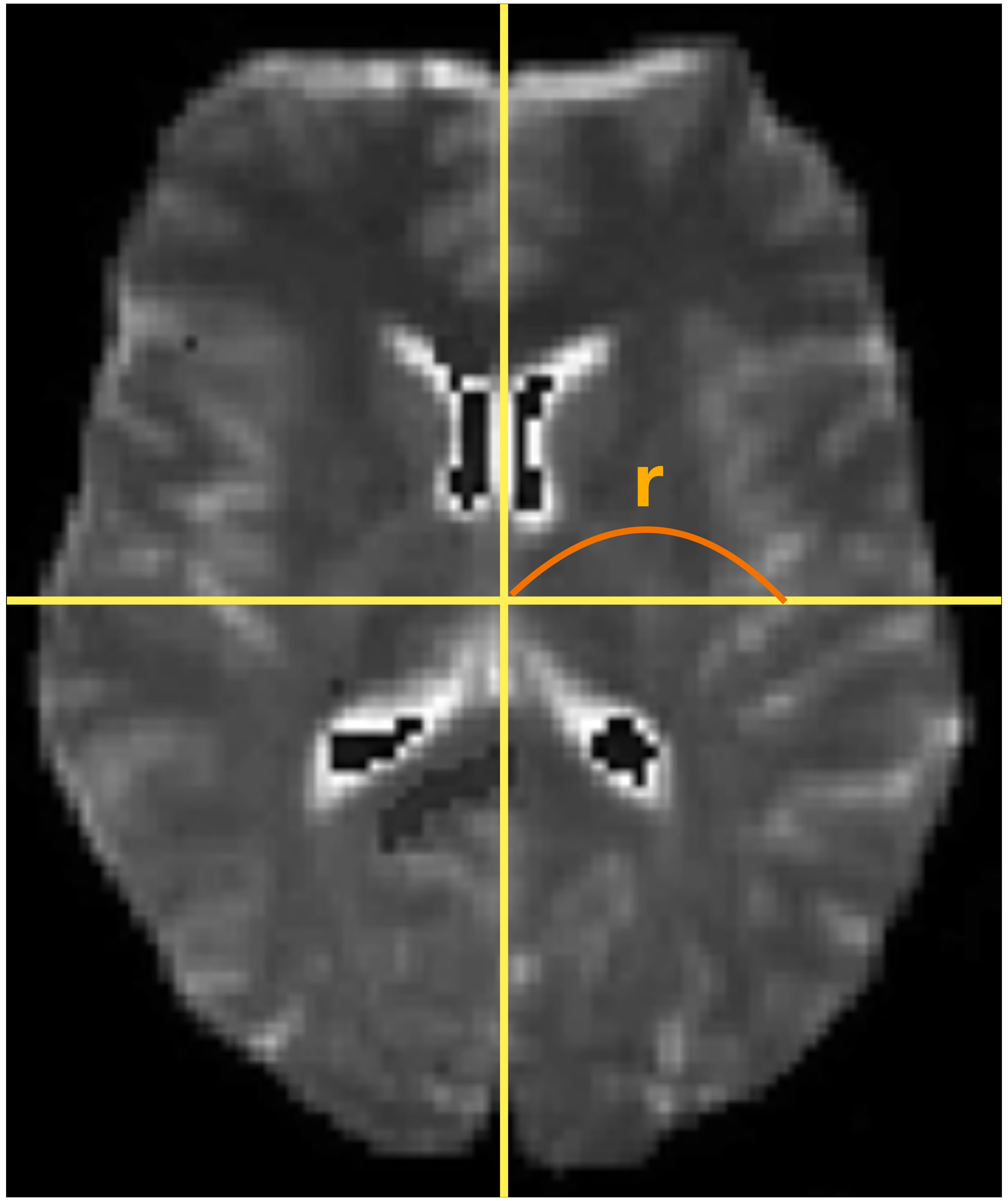

ICBM example: Age group difference in FA



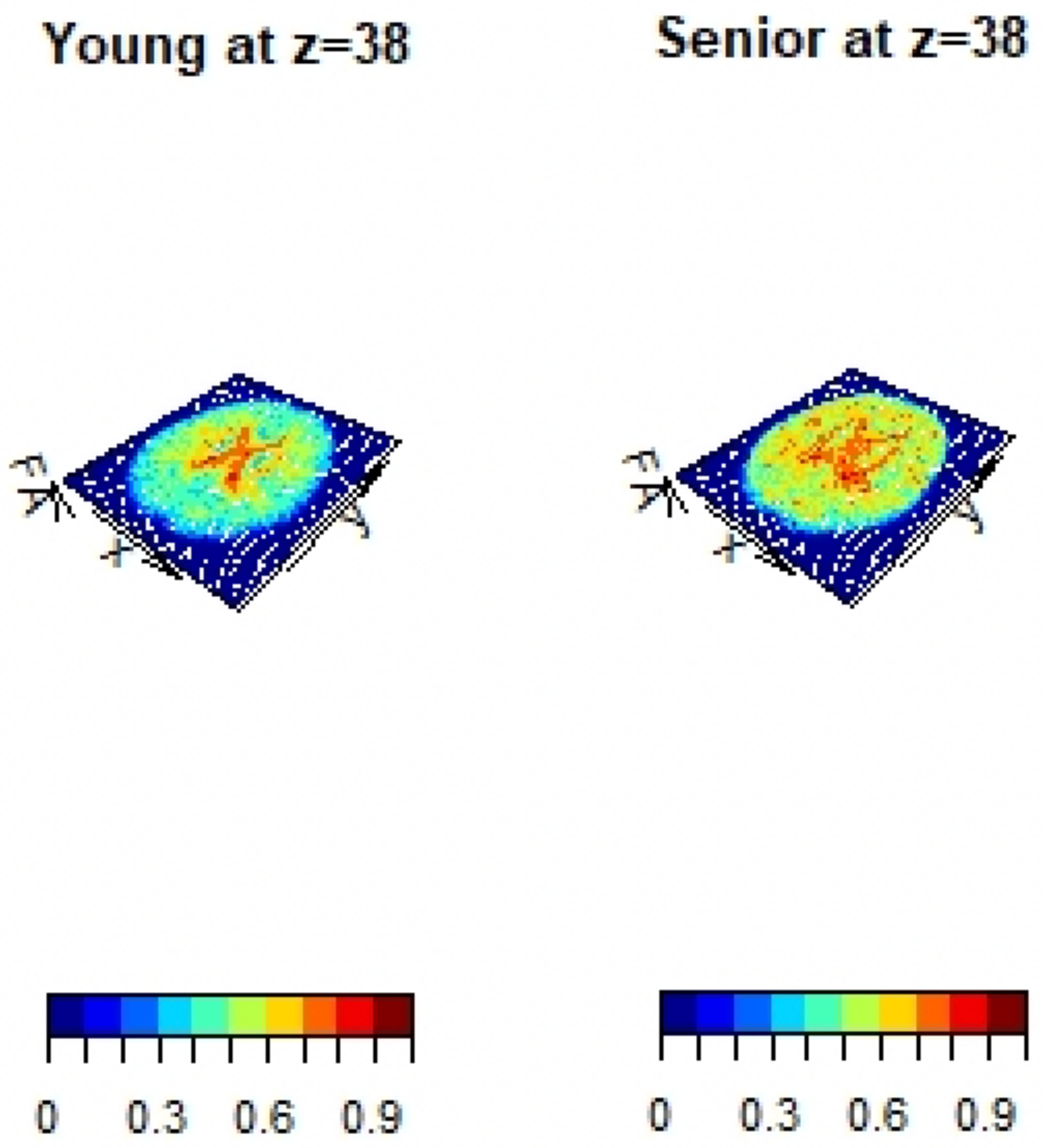
# Central Peripheral Deviation (CPD)



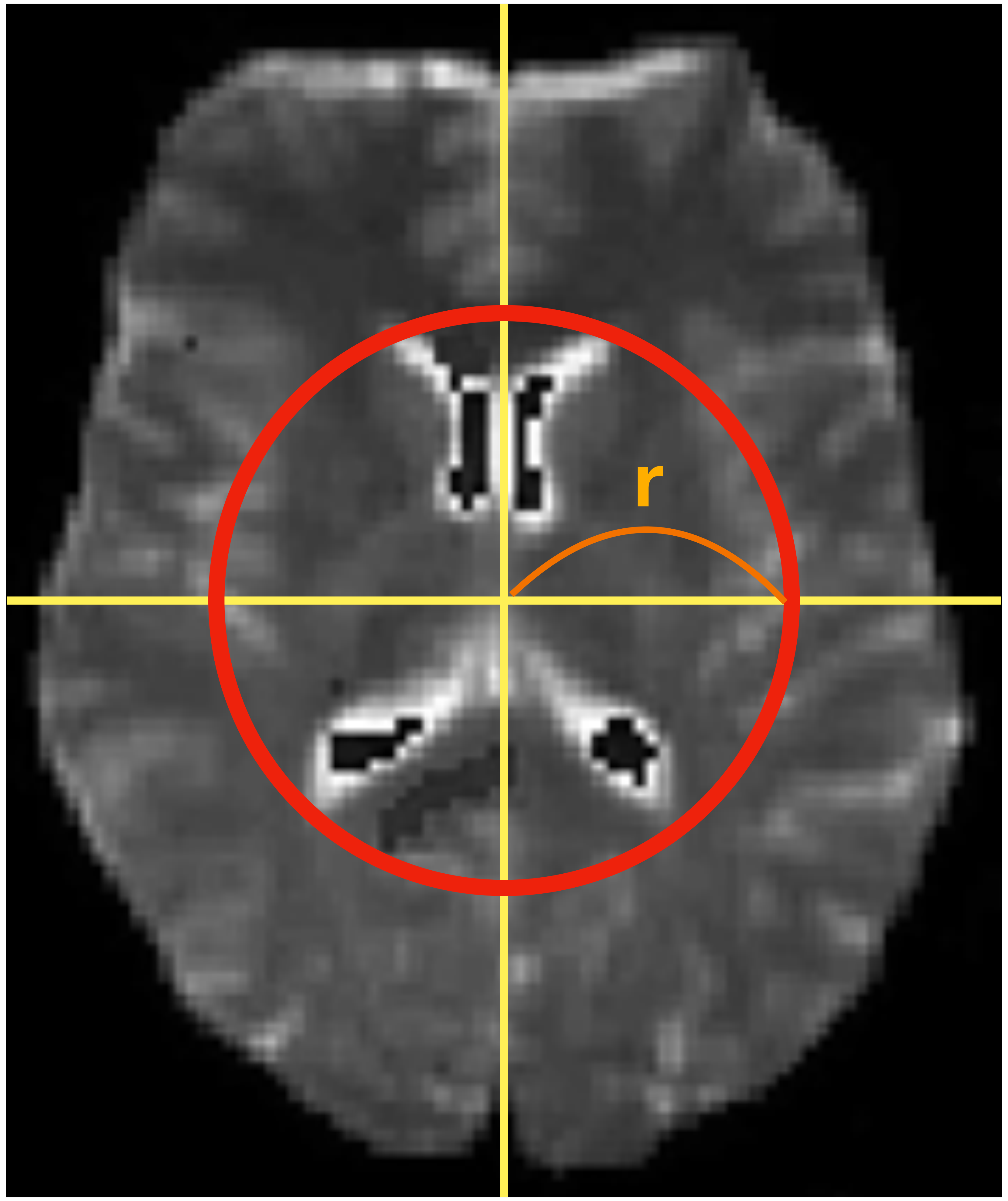

ICBM example: Age group difference in FA



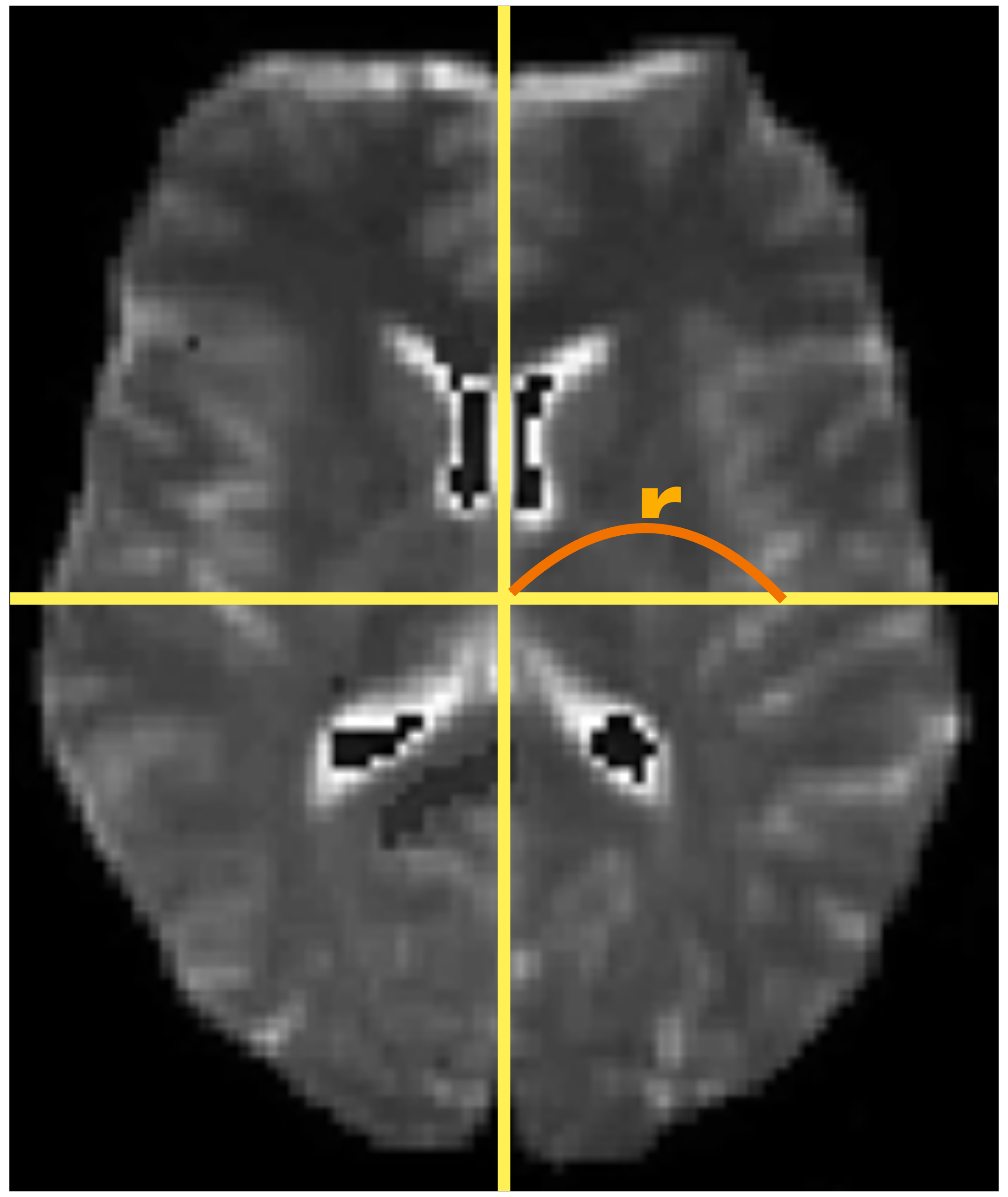
# Central Peripheral Deviation (CPD)




ICBM example: Age group difference in FA



# Definition of CPD



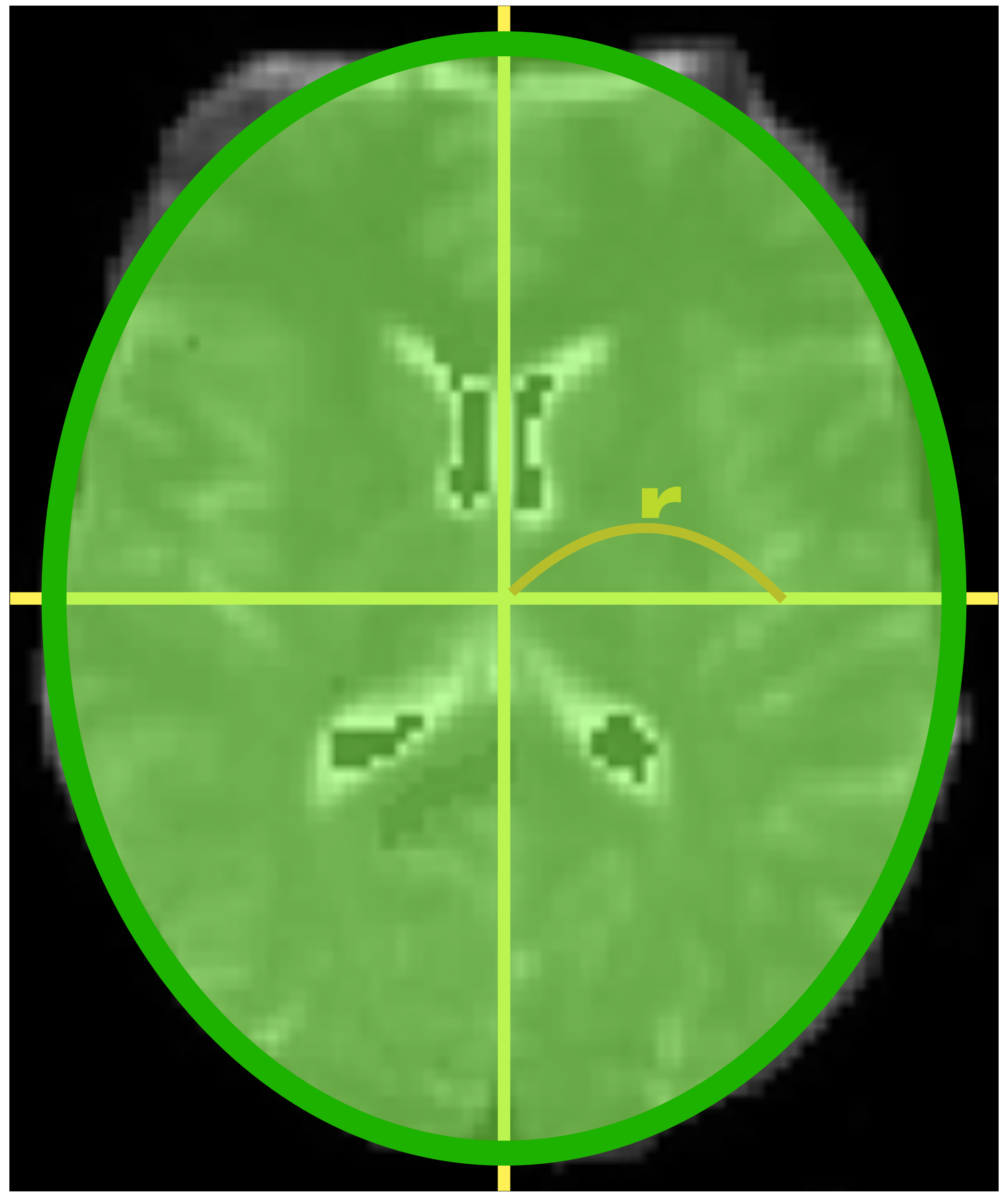

- $CPD = \mu_2 - (\mu_1 - \mu_2) = 2\mu_2 - \mu_1$
- $sCPD = \dfrac{\mu_2 - (\mu_1 - \mu_2)}{\sqrt{\varphi}} = \dfrac{2\mu_2 - \mu_1}{\sqrt{\varphi}}$

where $\mu_i = \dfrac{\int_{R_i} f(r,\theta)dA}{\int_{R_i} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_i} f(r,\theta)rdrd\theta}{\int_0^{2\pi}\int_0^{r_i} 1_{f(r,\theta)}rdrd\theta}$, $\varphi = \dfrac{\int_{R_1} \{f(r,\theta) - \mu_1\}^2 dA}{\int_{R_1} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_1} \{f(r,\theta) - \mu_1\}^2 rdrd\theta}{\int_0^{2\pi}\int_0 r_1 1_{f(r,\theta)}rdrd\theta}$, $i = 1{,}2$,

$r_2 < r_1 = \infty$ and $1_{f(r,\theta)} = \begin{cases} 1 & \text{if } f(r,\theta) > 0 \\ 0 & \text{if } f(r,\theta) = 0 \end{cases}$

# Definition of CPD

$\mu_1$: mean FA for entire brain




- $CPD = \mu_2 - (\mu_1 - \mu_2) = 2\mu_2 - \mu_1$
- $sCPD = \dfrac{\mu_2 - (\mu_1 - \mu_2)}{\sqrt{\varphi}} = \dfrac{2\mu_2 - \mu_1}{\sqrt{\varphi}}$

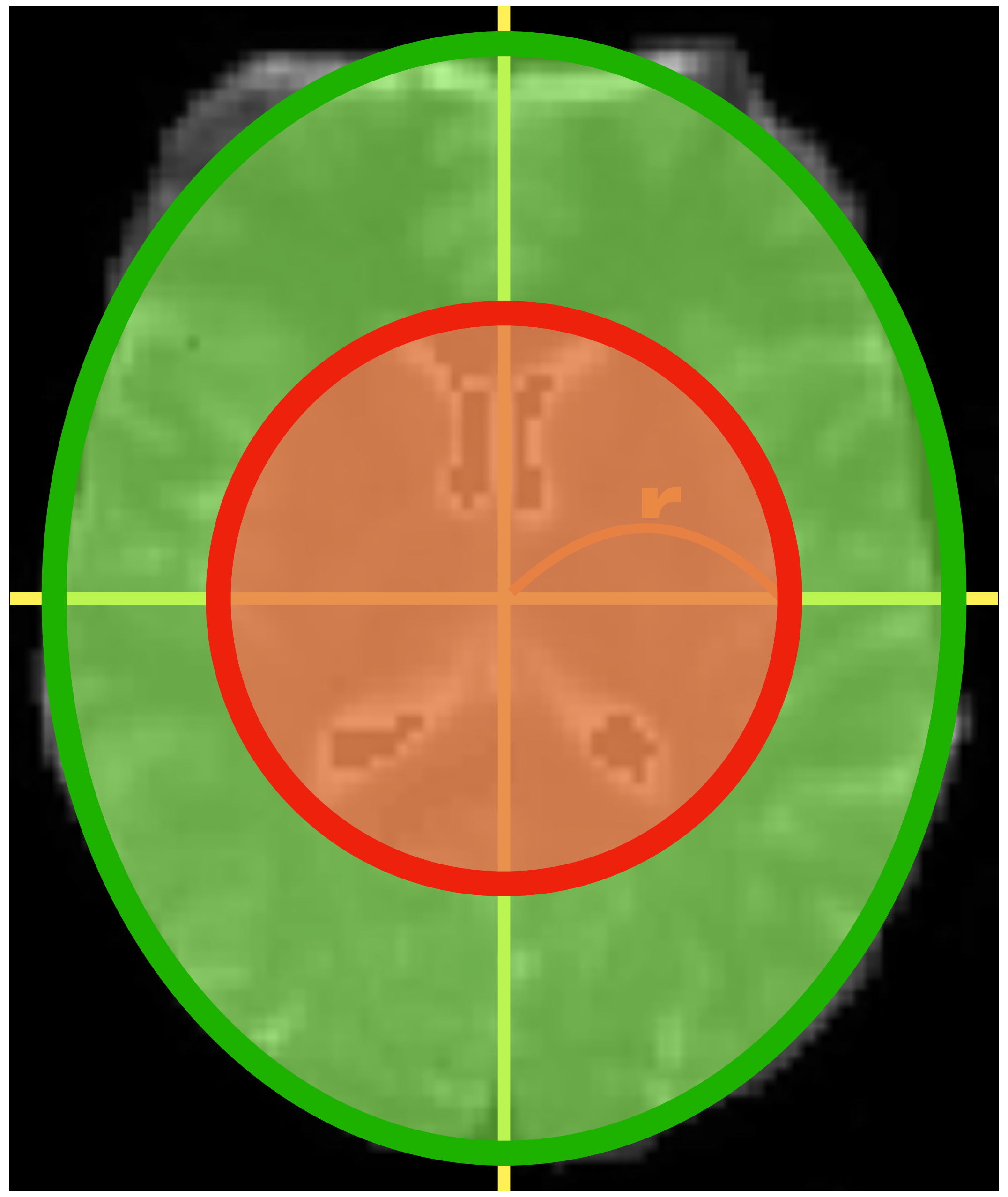
where $\mu_i = \dfrac{\int_{R_i} f(r,\theta)dA}{\int_{R_i} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_i} f(r,\theta)rdrd\theta}{\int_0^{2\pi}\int_0^{r_i} 1_{f(r,\theta)}rdrd\theta}$, $\varphi = \dfrac{\int_{R_1}\{f(r,\theta)-\mu_1\}^2dA}{\int_{R_1} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_1}\{f(r,\theta)-\mu_1\}^2rdrd\theta}{\int_0^{2\pi}\int_0 r_1 1_{f(r,\theta)}rdrd\theta}$, $i = 1{,}2$,

$r_2 < r_1 = \infty$ and $1_{f(r,\theta)} = \begin{cases} 1 & \text{if } f(r,\theta) > 0 \\ 0 & \text{if } f(r,\theta) = 0 \end{cases}$

# Definition of CPD

$\mu_1$: mean FA for entire brain

$\mu_2$: mean FA for inner circle




- $CPD = \mu_2 - (\mu_1 - \mu_2) = 2\mu_2 - \mu_1$
- $sCPD = \dfrac{\mu_2 - (\mu_1 - \mu_2)}{\sqrt{\varphi}} = \dfrac{2\mu_2 - \mu_1}{\sqrt{\varphi}}$

where $\mu_i = \dfrac{\int_{R_i} f(r,\theta)dA}{\int_{R_i} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_i} f(r,\theta)rdrd\theta}{\int_0^{2\pi}\int_0^{r_i} 1_{f(r,\theta)}rdrd\theta}$, $\varphi = \dfrac{\int_{R_1}\{f(r,\theta)-\mu_1\}^2 dA}{\int_{R_1} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_1}\{f(r,\theta)-\mu_1\}^2 rdrd\theta}{\int_0^{2\pi}\int_0 r_1 1_{f(r,\theta)}rdrd\theta}$, $i = 1,2,$

$r_2 < r_1 = \infty$ and $1_{f(r,\theta)} = \begin{cases} 1 & \text{if } f(r,\theta) > 0 \\ 0 & \text{if } f(r,\theta) = 0 \end{cases}$

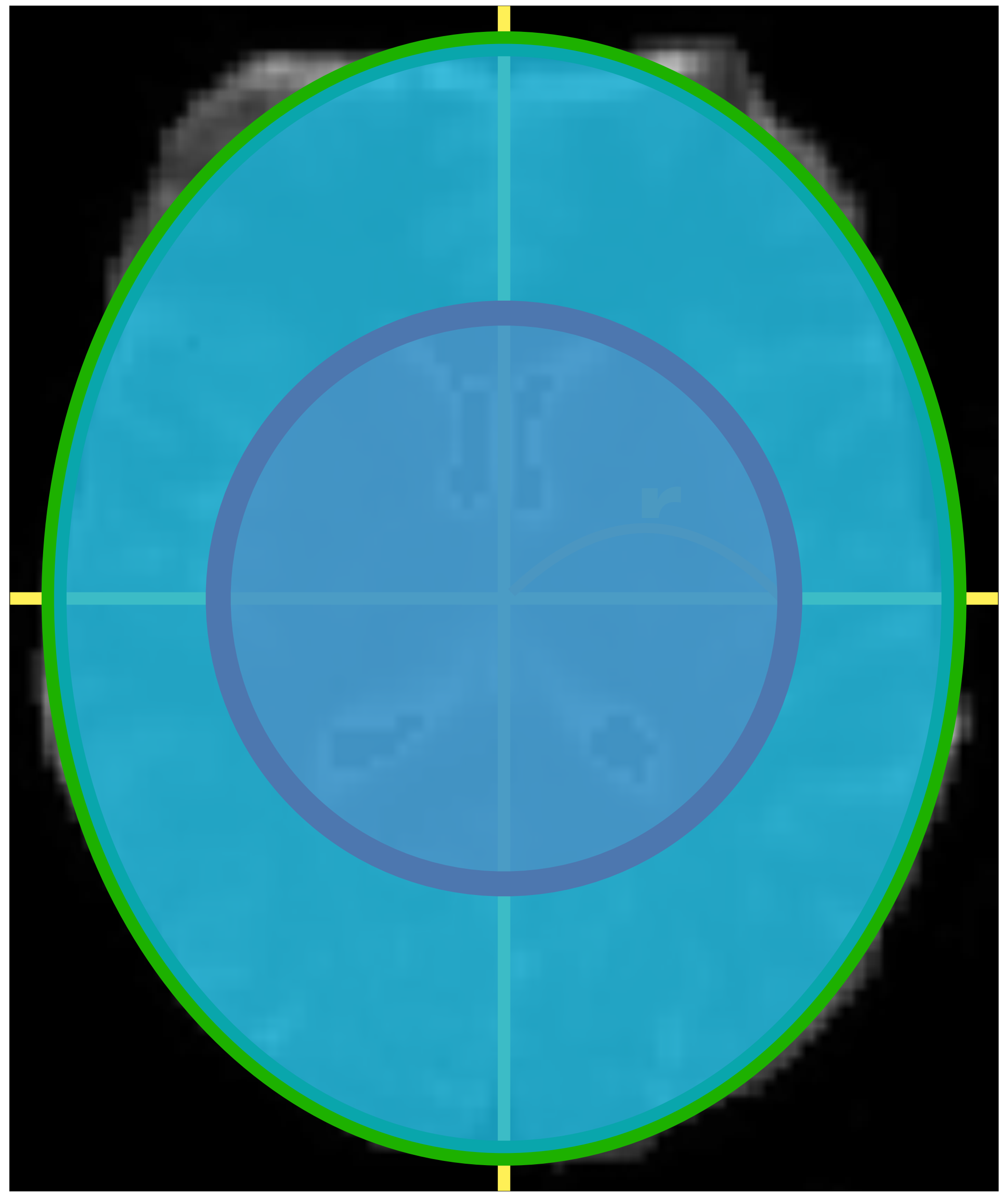
# Definition of CPD

- $CPD = \mu_2 - (\mu_1 - \mu_2) = 2\mu_2 - \mu_1$
- $sCPD = \dfrac{\mu_2 - (\mu_1 - \mu_2)}{\sqrt{\varphi}} = \dfrac{2\mu_2 - \mu_1}{\sqrt{\varphi}}$

$\mu_1$: mean FA for entire brain

$\mu_2$: mean FA for inner circle

$\varphi$: variance FA for entire brain

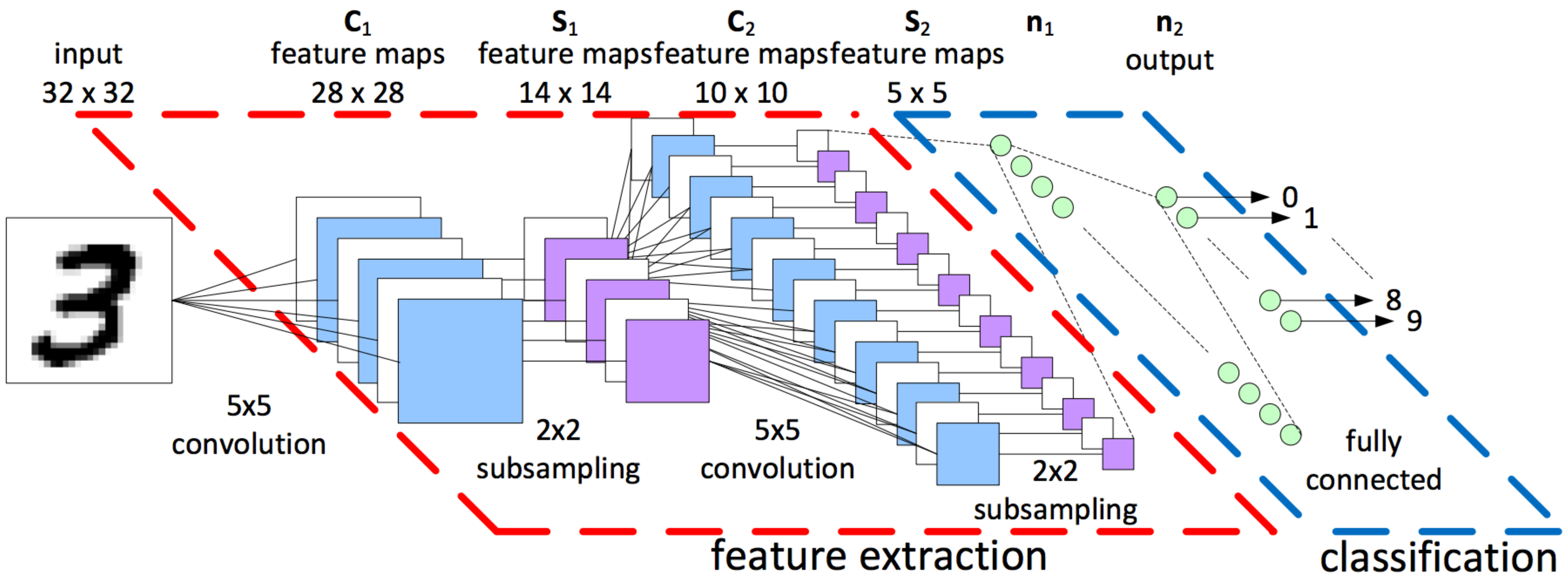


where $\mu_i = \dfrac{\int_{R_i} f(r,\theta)dA}{\int_{R_i} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_i} f(r,\theta)rdrd\theta}{\int_0^{2\pi}\int_0^{r_i} 1_{f(r,\theta)}rdrd\theta}, \varphi = \dfrac{\int_{R_1}\{f(r,\theta)-\mu_1\}^2 dA}{\int_{R_1} 1_{f(r,\theta)}dA} = \dfrac{\int_0^{2\pi}\int_0^{r_1}\{f(r,\theta)-\mu_1\}^2 rdrd\theta}{\int_0^{2\pi}\int_0 r_1 1_{f(r,\theta)}rdrd\theta}, i = 1,2,$

$r_2 < r_1 = \infty$ and $1_{f(r,\theta)} = \begin{cases} 1 & \text{if } f(r,\theta) > 0 \\ 0 & \text{if } f(r,\theta) = 0 \end{cases}$

# Convolutional Neural Network (CNN)

- Hierarchical layered representation learning
- A 2D representation via “back-propagation”



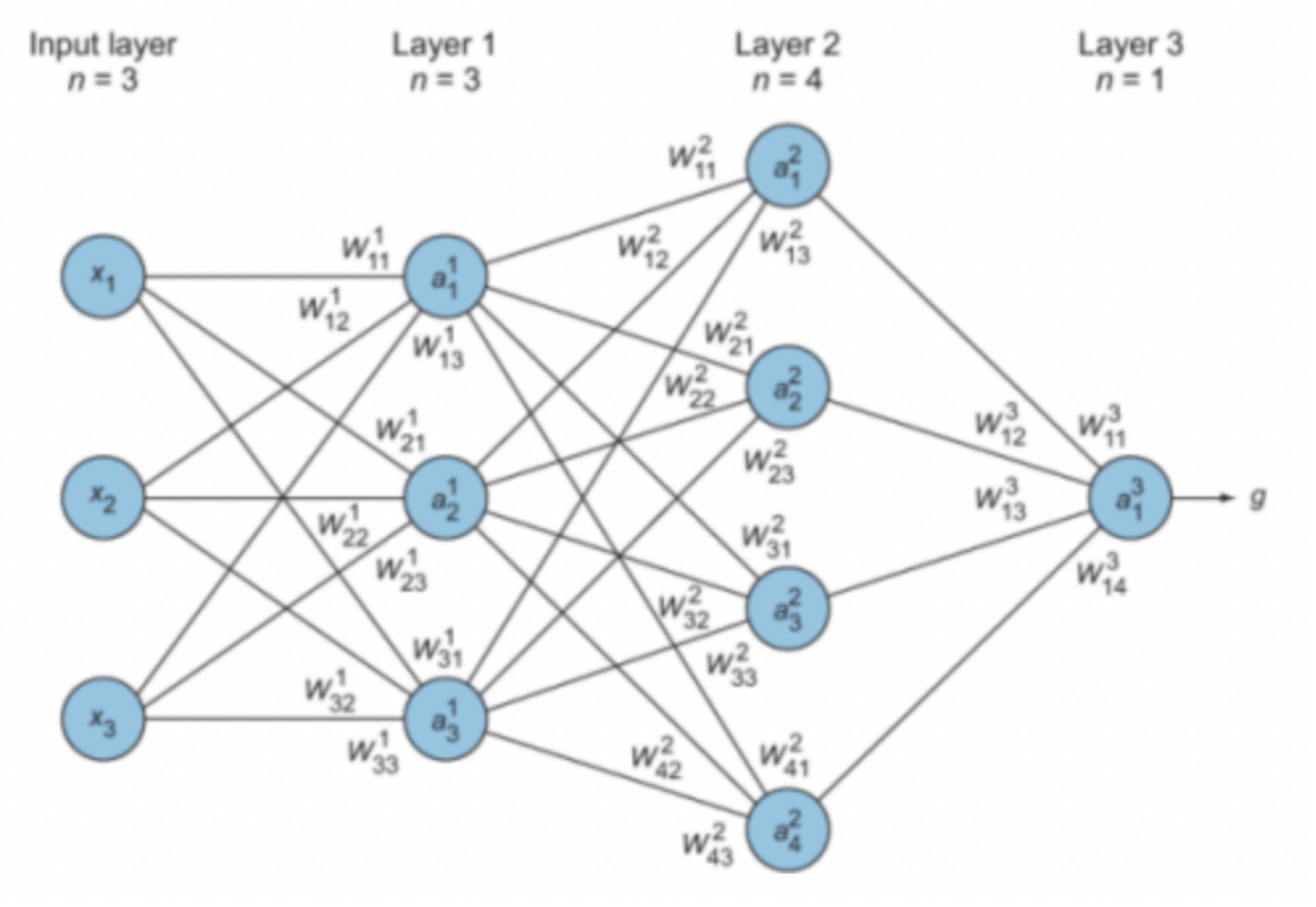

Credit: kaggle.com

# Convolutional Neural Network (CNN)

- Find a set of values for the weights of all layers in a network associated with targets

- Adjust the value of the weights in direction of the lower cost (or loss) by an objective function, e.g., Binary Cross-Entropy

$$Loss = -\frac{1}{n}\sum_{1}^{n}\left\{y_i \times \log\big(p(y_i)\big) + (1-y_i)\times\log\big(1-p(y_i)\big)\right\}$$




Multi-Layer Perceptron (MLP)

Credit: Algendy 2020.

# Applying CPD to CNN architecture

- Recent work on imaging data
  - Application of CNN to automatic *segmentation* of breast lesion in Ultra Sound images (Costa et al., 2019)

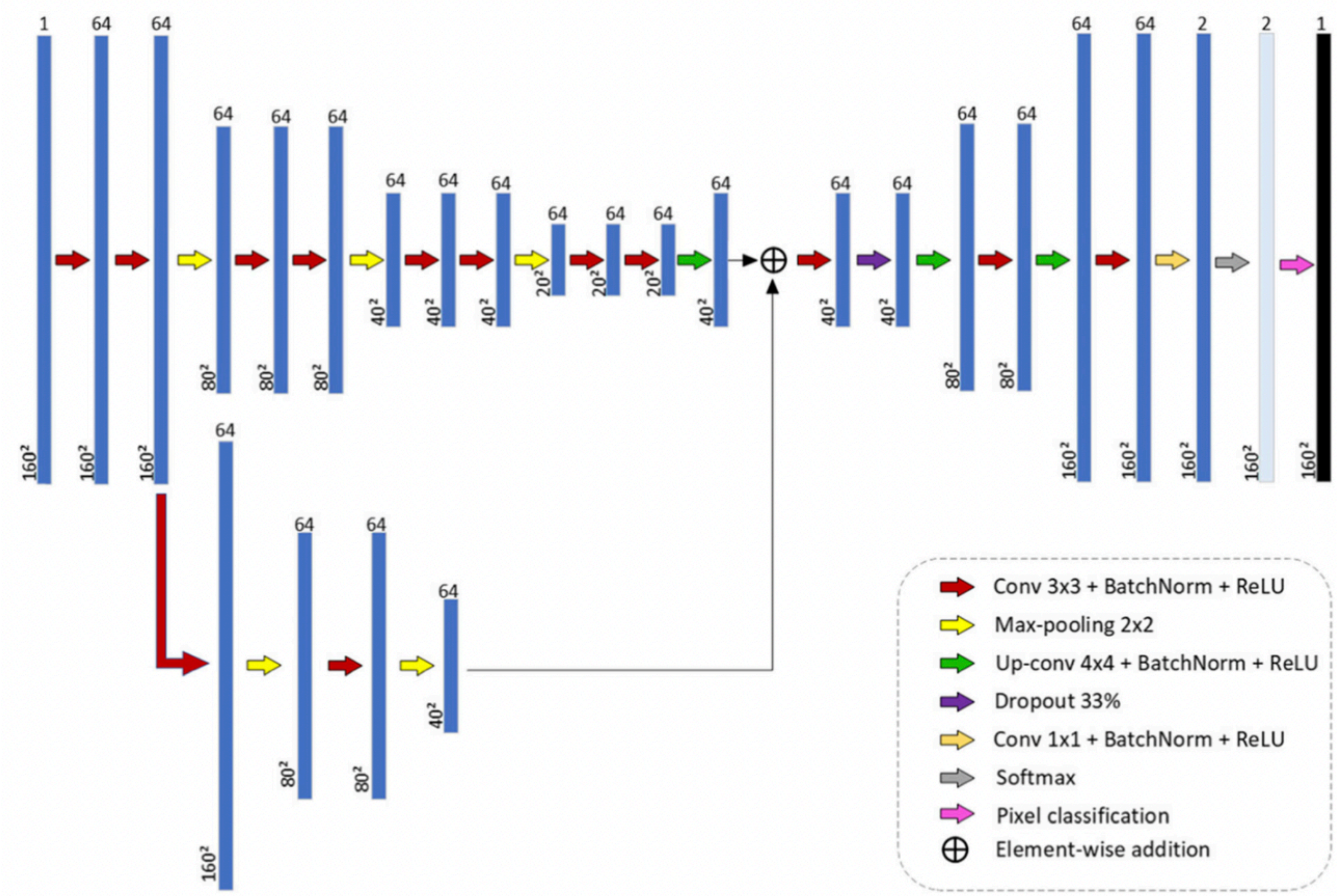



**Fig. 3** CNN2 architecture: first DAG architecture proposed for breast lesion segmentation in US image

# Proposed CNN architecture

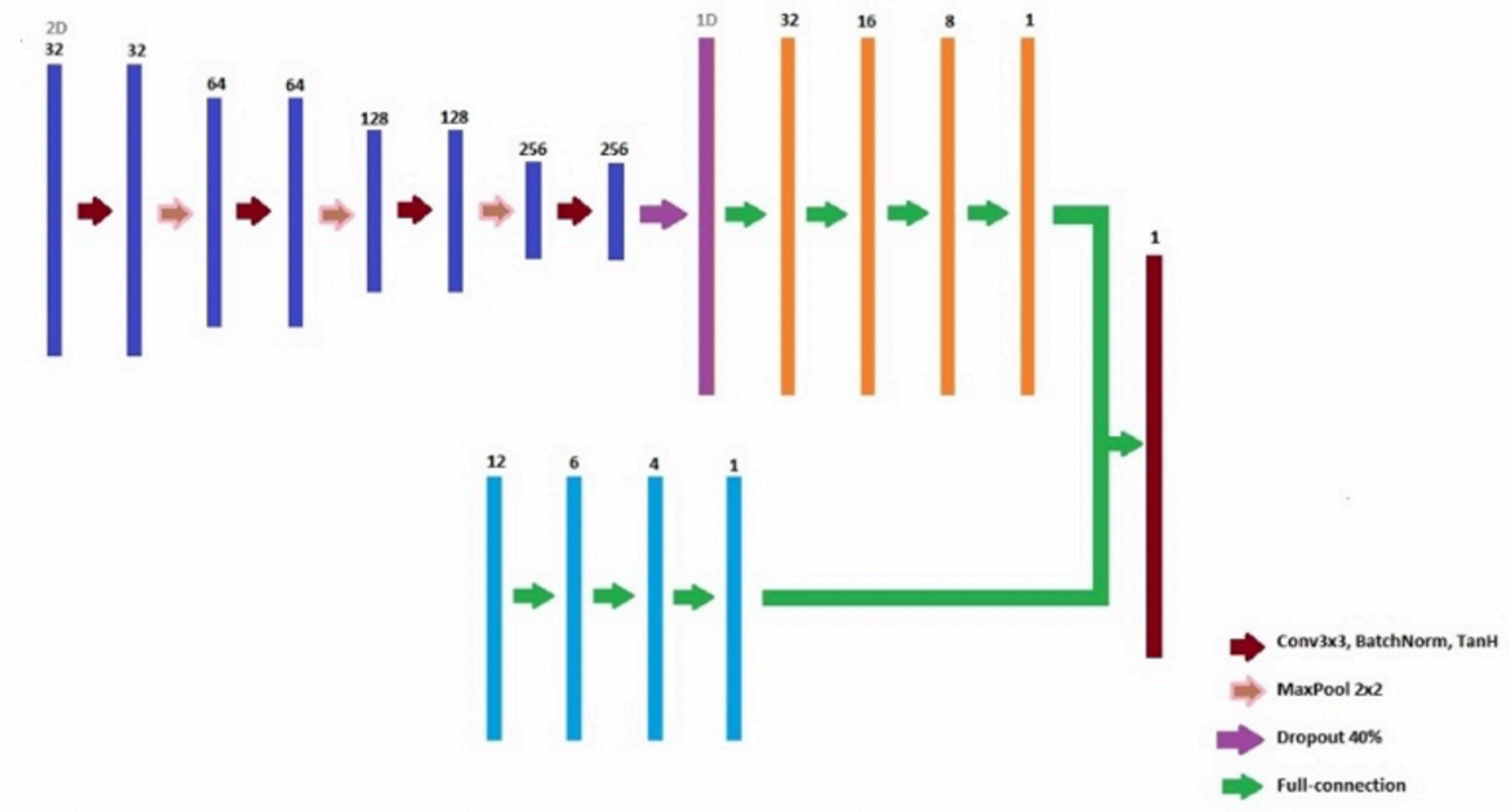



Figure 4. The proposed network architecture using CPD. The upper layers describe the CNN and the lower layers describe the MLP with 4 layers. The smaller the length of bar indicates the smaller size of the images. The number on top of each bar represents the number of nodes at the layer. The arrows with different colors represent different procedure as shown in the legend.

# Proposed CNN architecture

- Key points of CNN+MLP with CPD

1. Adopt CNN: translation-invariant, spatial hierarchies for 2D images
2. **Good features** let you solve a problem with far less data
3. Training a **simple, well regularized** CNN model on a **very small data set** can potentially suffice and yield reasonable results
   - Simple: a small number of layers
   - Well-regularized: an optimized L2-norm $\lambda$ value

# Contents

# Example DTI studies

- Study 1: Maryland MagNeTS prospective study from FITBIR
  - Traumatic Brain Injury (TBI) patients with baseline brain images
  - Access only DTI images (n=82) and estimate tensor information
- Study 2: International Consortium for Brain Mapping (ICBM) from LONI
  - Normal subjects' brain images for developing a MNI atlas template
  - Access only DTI images (n=192) and estimate tensor information

# Training the model

- Input (X): DTI images and CPD
- Output (Y): Binary age group (Age $\leq 52$ or $> 52$)
- Split data into training data and test data at ratio 80:20

# Improved test accuracy on proposed model

## Results from MagNeTS data

Table 3. The models with best test accuracies for the CNN method alone and the CNN + MLP method using the MagNeTS data.

| Model | Input data | MLP design | Radius $r$ | $\lambda$ values $(\lambda_{mlp}, \lambda_{cnn})$ | Training loss | Training accuracy | Test loss | Test accuracy |
|---|---|---|---|---|---|---|---|---|
| CNN | Image | | | $(NA, 3)$ | 293.1302 | 0.7005 | 292.5994 | 0.6250 |
| CNN+MLP | Image, unscaled CPDs | 2 layers | 5 | $(0.01, 0.5)$ | 51.3255 | 0.6973 | 51.5627 | 0.7083 |
| | | | 15 | $(0.1, 0.5)$ | 52.6860 | 0.7582 | 51.3741 | 0.7083 |
| | | | 5, 15 | $(0.01, 0.5)$ | 51.3449 | 0.7659 | 50.8401 | 0.7083 |
| | | 4 layers | 5 | $(1, 0.5)$ | 70.7909 | 0.7183 | 70.9441 | 0.7083 |
| | | | 15 | $(0.1, 0.5)$ | 53.2396 | 0.7872 | 52.6477 | 0.7500 |
| | | | 5, 15 | $(0.01, 0.5)$ | 52.0297 | 0.8125 | 51.9336 | 0.7500 |
| | Image, scaled CPDs | 2 layers | 5 | $(1, 0.5)$ | 0.7184 | 0.7414 | 0.7625 | 0.7083 |
| | | | 15 | $(0.05, 0.5)$ | 0.8342 | 0.7586 | 0.8405 | 0.7083 |
| | | | 5, 15 | $(0.1, 0.5)$ | 1.6615 | 0.7241 | 1.6553 | 0.7500 |
| | | 4 layers | 5 | $(0.01, 0.5)$ | 1.4807 | 0.7414 | 1.5009 | 0.7083 |
| | | | 15 | $(1, 0.5)$ | 1.4562 | 0.7414 | 1.4908 | 0.7083 |
| | | | 5, 15 | $(0.1, 0.5)$ | 2.4740 | 0.7069 | 2.4500 | 0.7500 |

# Improved test accuracy on proposed model

## Results from ICBM data

Table 5. The models with best test accuracies for the CNN method alone and the CNN + MLP method using the ICBM data. We train the CNN + MLP model with both $CPD_5$ and $CPD_{15}$. Complete data rate indicates the threshold percentage of available FA values inside the basis circle from the center image. For example, 70% indicates the images where 70% or greater of FA values are available inside circles at both radii.

| Model | Input data | Complete data rate | $\lambda$ values $(\lambda_{mlp}, \lambda_{cnn})$ | Training loss | Training accuracy | Test loss | Test accuracy |
|---|---|---|---|---|---|---|---|
| CNN | Image | | $(NA, 0.1)$ | 11.3143 | 0.7188 | 12.3710 | 0.6379 |
| CNN+MLP | Image, unscaled CPDs | 70% | $(3, 0.1)$ | 56.6895 | 0.7031 | 56.7059 | 0.6897 |
| | | No restriction | $(0.1, 0.1)$ | 13.1572 | 0.6493 | 15.6420 | 0.6379 |
| | Image, scaled CPDs | 70% | $(3, 0.1)$ | 50.2584 | 0.7031 | 49.8063 | 0.6897 |
| | | No restriction | $(0.01, 0.1)$ | 11.2429 | 0.7463 | 12.4016 | 0.6724 |

# Contents

# New paradigm for DTI study

- Our consideration of algebraic variants for brain's characteristic may complement information not revealed through CNN's feature selection

# New paradigm for DTI study

- Our consideration of algebraic variants for brain's characteristic may complement information not revealed through CNN's feature selection

| Advantages | Disadvantages |
|---|---|
| A. No experts' opinion to search for regions | A. Need more physiologically relevant representation |
| B. CPD may not require very crisp images | B. Data size is not enough for generalization |
| C. CPD may complement information not revealed by CNN's feature selection | |

# New paradigm for DTI study

- Future work
  1. Extend it to multi-centric peripheral deviation
  2. Apply it to 3D in order to verify justification of 2D approach
  3. Find the best criteria for $\lambda -$ change in each layer
  4. Use other diffusion coefficients such as Mean Diffusivity, Radial Diffusivity, Axial Diffusivity

# Thank you