Bayesian Image Quality Transfer with CNNs: Exploring Uncertainty in dMRI Super-Resolution

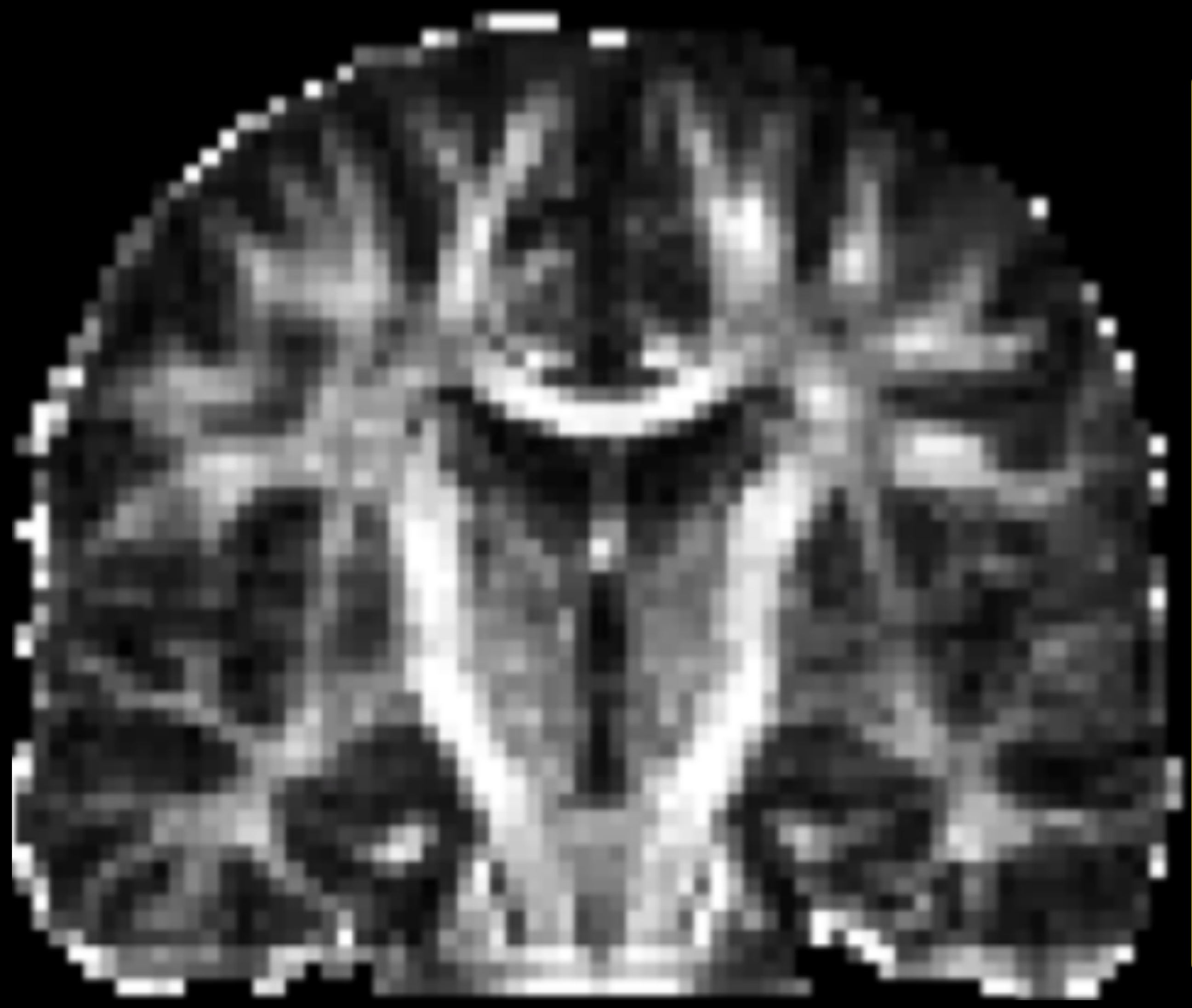
Ryutaro Tanno, Daniel Worrall, Aurobrata Ghosh, Enrico Kaden, Stamatios N. Sotiropoulos, Antonio Criminisi, Daniel C. Alexander

20th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI 2017)
September 2017, Quebec

What is Image Quality Transfer?

UCL

humanconnectome.org
[Sotiropoulos et al. NIMG 2013]



Clinical scanners

- Low spatial resolution and SNR
- Time and cost pressure
- So, limited quality of subsequent analysis

Special scanners

- High spatial resolution and SNR
- Longer acquisition time
- Expensive

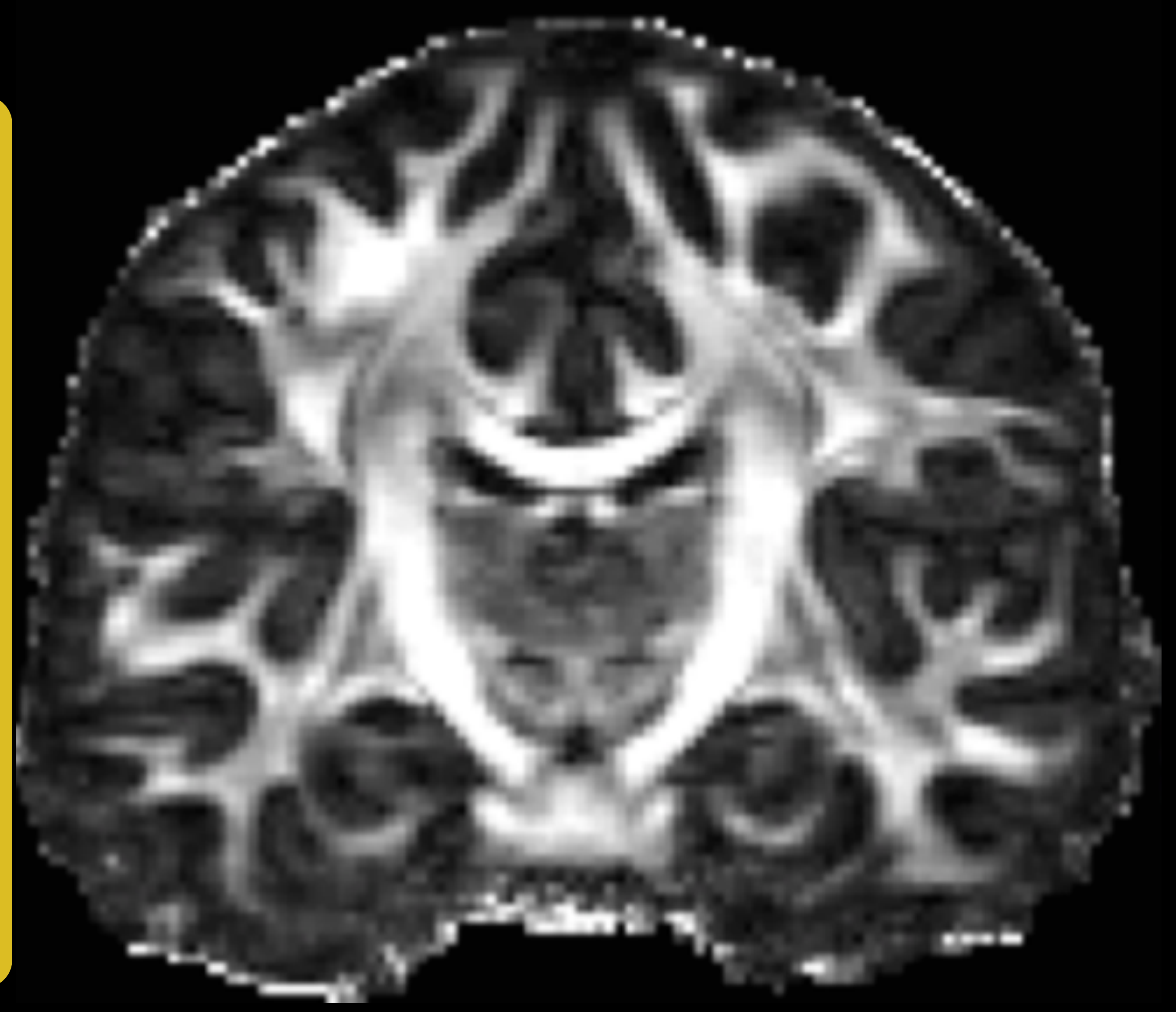


Image Quality Transfer (IQT)

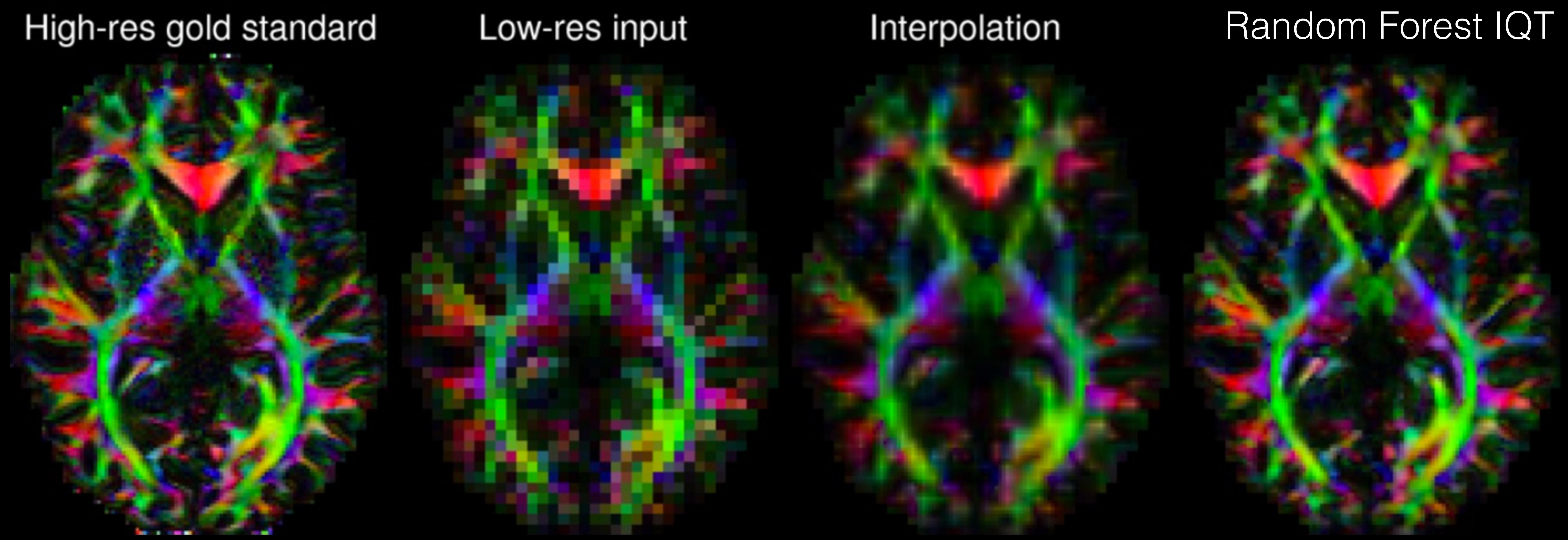
[Alexander et al. NIMG'17, Tanno et al. MICCAI'16]

- **Machine learning for quality enhancement**
- **Propagating information in high quality data from special scanners.**

What is Image Quality Transfer?

UCL

- **Random forest IQT** [Alexander et al. MICCAI'14, NIMG'17]
  1. super-resolution of DTI/MAP-MRI and downstream tractography
  2. estimation of advanced microstructure contrasts (e.g. NODDI, SMT) from DTIs.

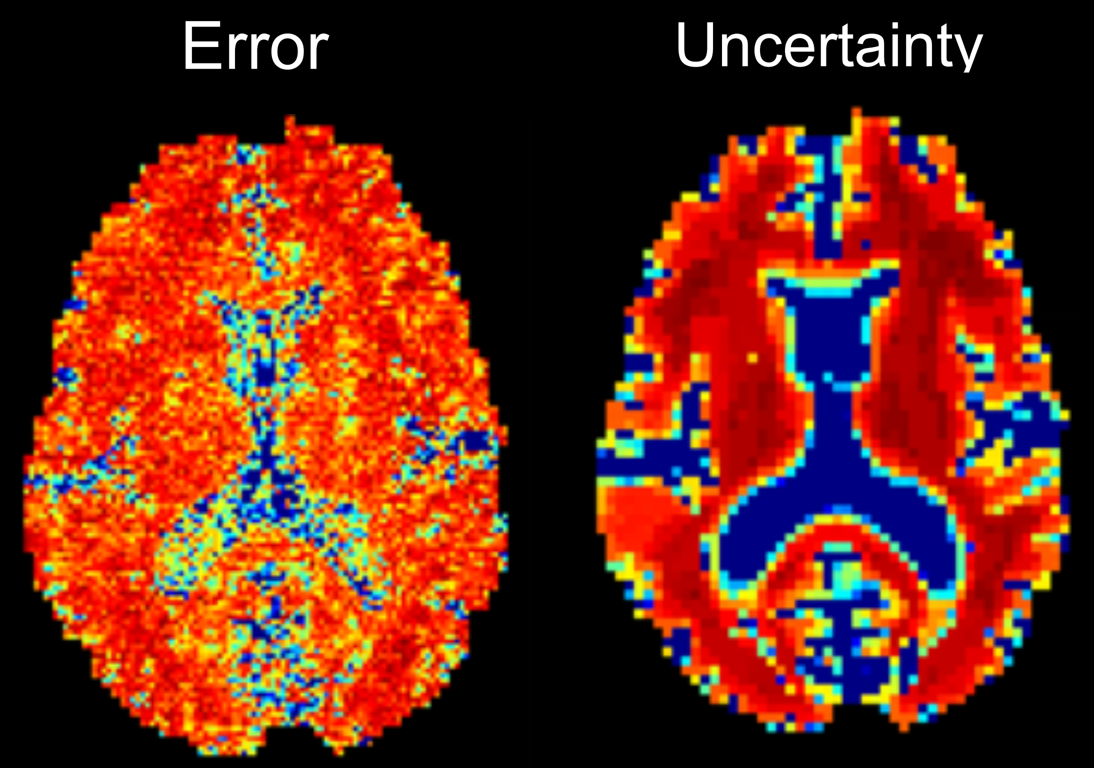


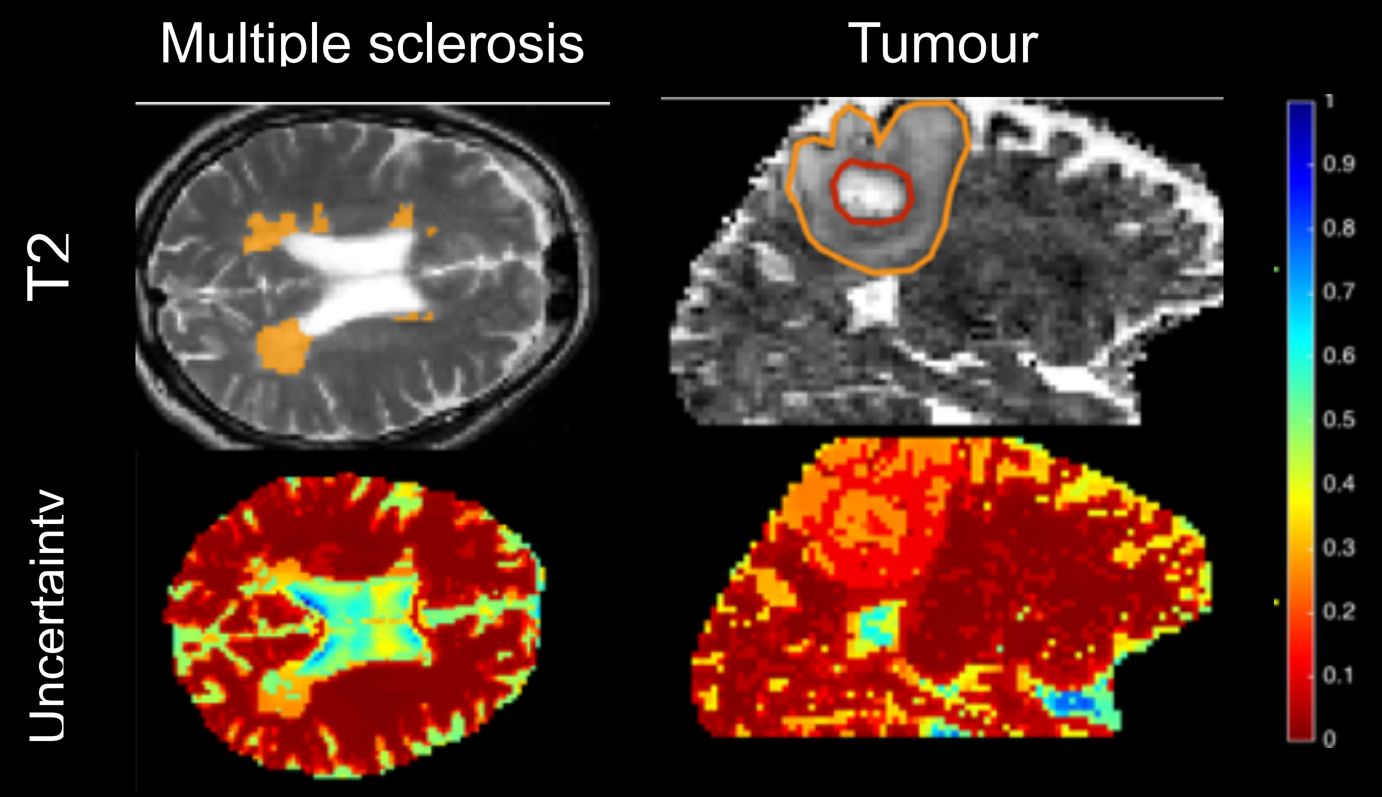
What is Image Quality Transfer?

- **Random forest IQT** [Alexander et al. MICCAI'14, NIMG'17]
  1. super-resolution of DTI/MAP-MRI and downstream tractography
  2. estimation of advanced microstructure contrasts (e.g. NODDI, SMT) from DTIs.

> **LIMITATION:** no indication of **uncertainty** in predicted enhanced image

- **Bayesian IQT** [Tanno et al. MICCAI'16]
  1. proposed a locally Bayesian variant of random forests
  2. estimate of predictive uncertainty which highly correlates with accuracy





Goals

UCL

- Goal: Devise a **deep learning** implementation of **Bayesian IQT**

- Promising applications of deep learning to related problems:

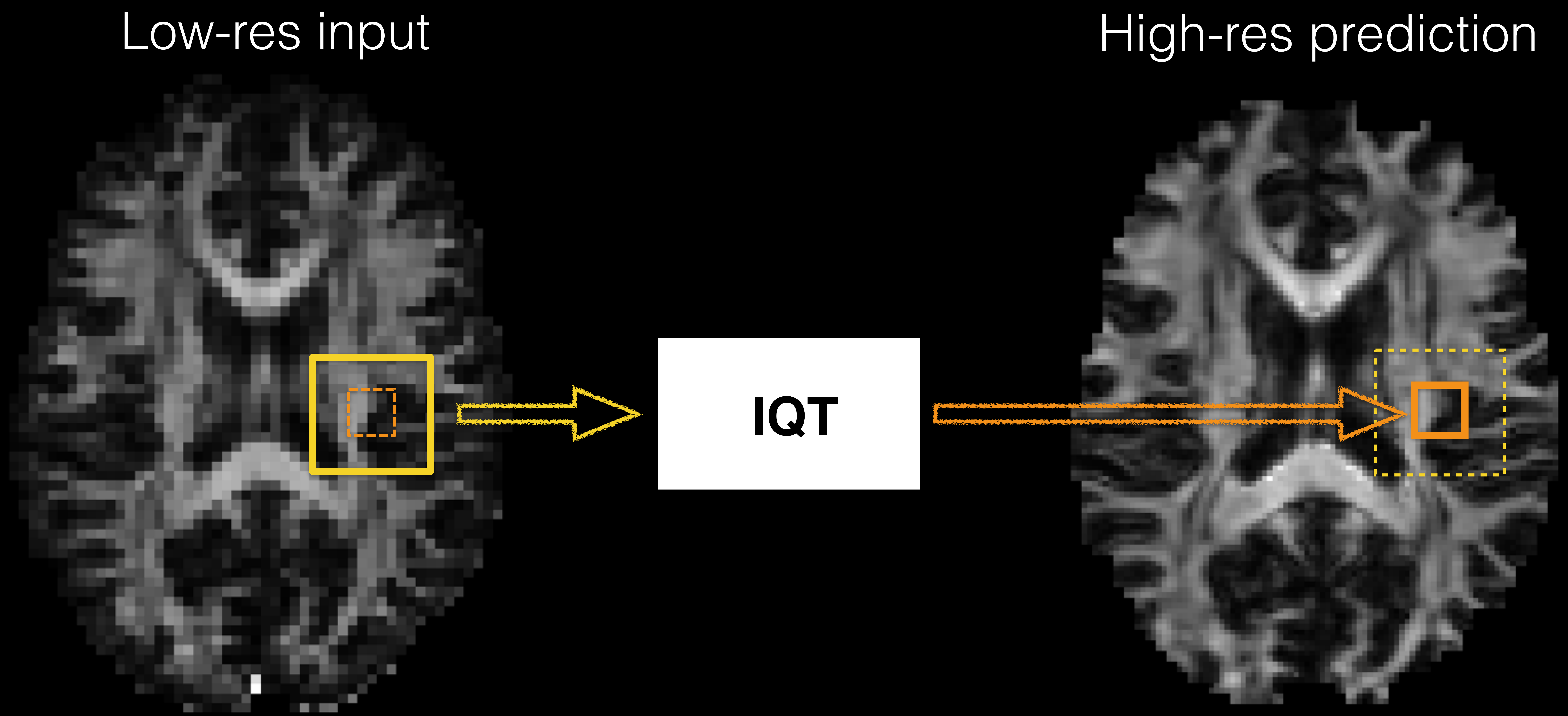
    1. super-resolution, e.g. cardiac MRI [Oktay et al. MICCAI'16]
    2. contrast transfer, e.g. predicting 7T contrast from 3T image [Bahrami. MICCAI'17]
    3. sparse MR reconstruction: [Schlemper et al. IPMI'17, Mardani et al. 2017]
    4. denoising: [Gondara et al. 2016, Jifara et al. 2017]
    5. dealiasing, motion correction: [Yu et al. 2017]

- This work aims to:

    1. **test performance benefits of deep learning to IQT**
    2. **explore ways to estimate different types of uncertainty in quality enhancement**

- Demonstrate in **super-resolution** of diffusion MR

Super-resolution as patch-regression

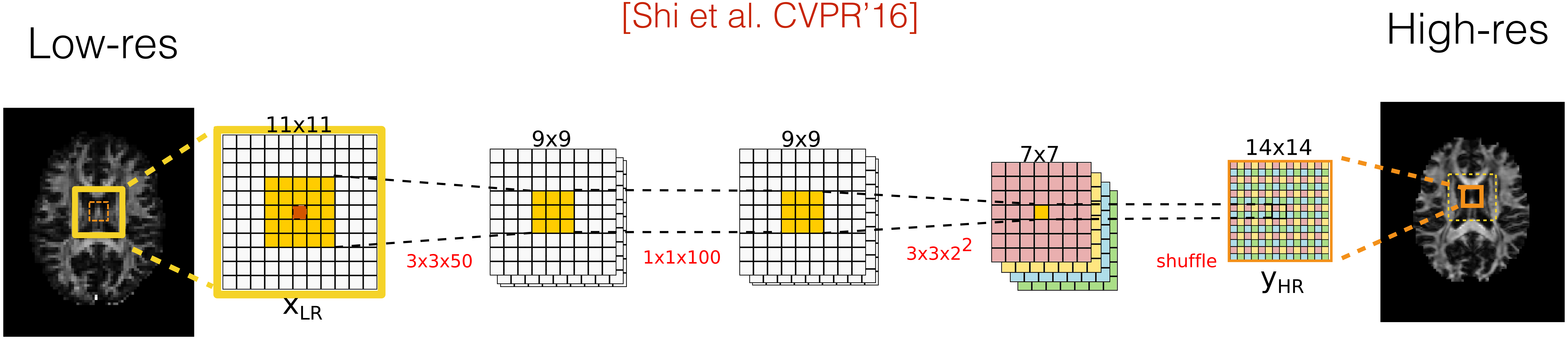


2D Illustration

Baseline 3D super-resolution network

2D Illustration

ESPCN = Efficient Subpixel Convolutional Network,
[Shi et al. CVPR'16]

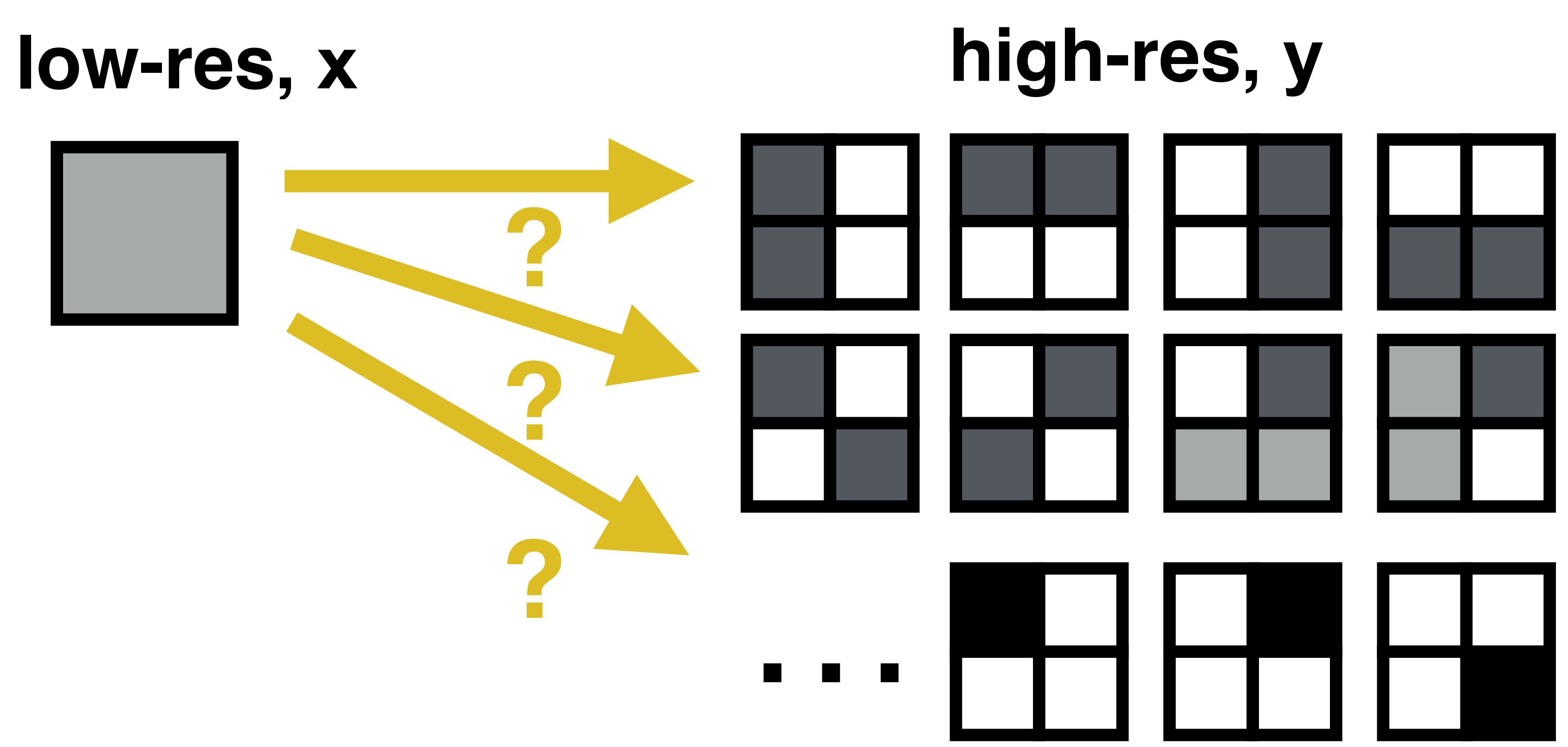


- **last conv. + shuffle = deconv. = learned interpolation**
- **Trained to minimise average pixel-wise MSE**
- **Two advances:**
  - **(I). 3D Extension of ESPCN**
  - **(II). Probabilistic Extensions for Modelling Uncertainty**

Uncertainty Modelling

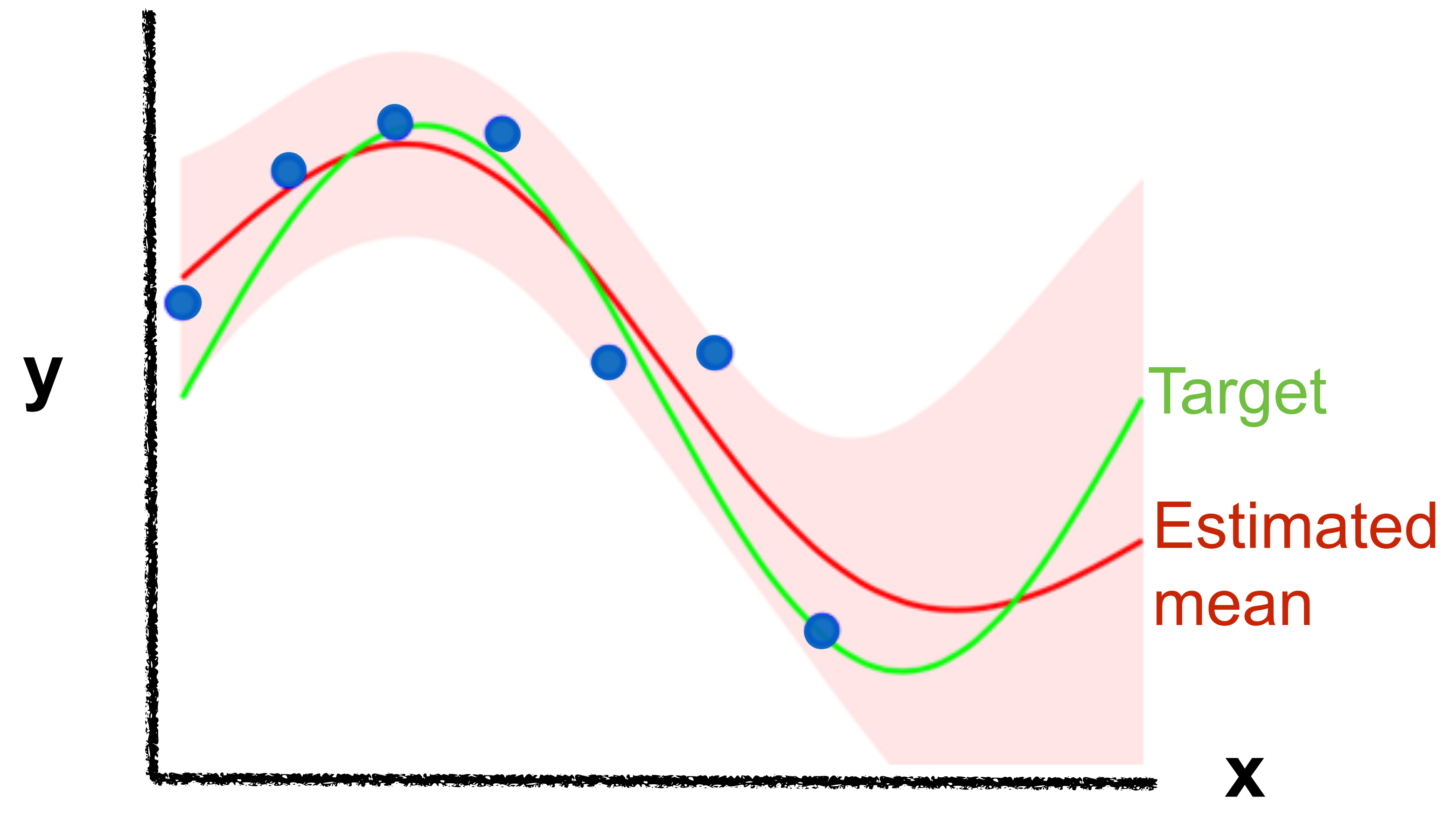
- Model **two components** of uncertainty in super-resolution

(i) **Intrinsic uncertainty**



- inherent ambiguity in the problem e.g. one-to-many nature of super-resolution mapping.
- cannot be reduced even with **infinite** data.
-

(ii) **Parameter uncertainty**



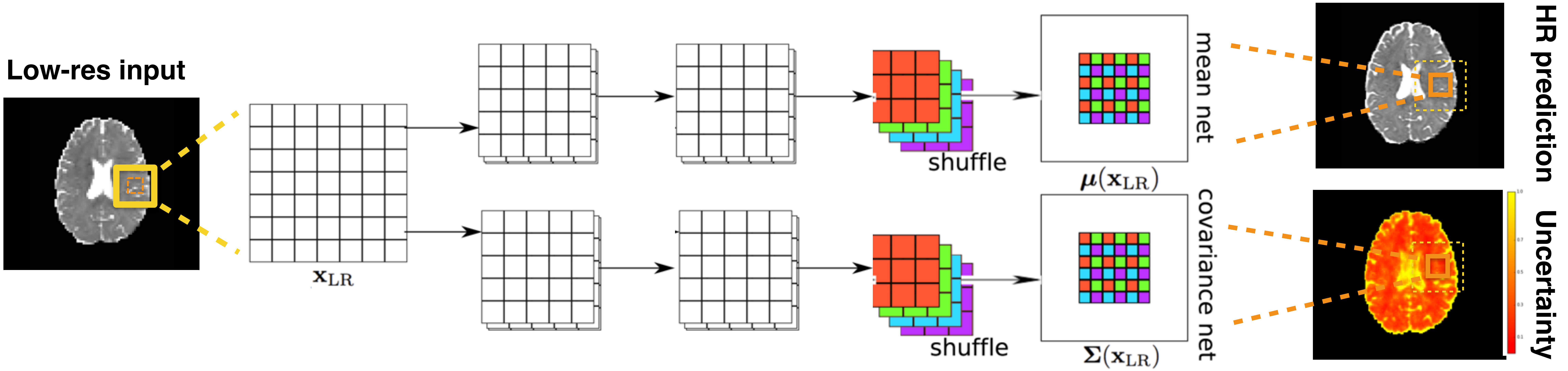
- ambiguity in the choice of “best” model parameters.
- can be explained away with **infinite** data

➡ More generalisable prediction

➡ Quantification of reliability (e.g., confidence interval)

Intrinsic Uncertainty:
Heteroscedastic noise model

- **Model intrinsic uncertainty as a spatially varying multivariate Gaussian distribution [Nix et al. 1994]**



- **Dual network architecture:** **two separate 3D-ESPCNS** to estimate the **mean** and **covariance** of the Gaussian likelihood.
- **Jointly optimised** to minimise the negative log likelihood on observations $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_i^{|\mathcal{D}|}$

$$\mathcal{L}_\theta(\mathcal{D}) = -\sum_{(\mathbf{x}_i,\mathbf{y}_i)\in\mathcal{D}} \log \mathcal{N}(\mathbf{y}_i; \mu_{\theta_1}(\mathbf{x}_i), \Sigma_{\theta_2}(\mathbf{x}_i))$$

- **No parameter sharing** between mean and covariance networks

Parameter Uncertainty:
Variational Dropout

- Previous methods rely on a **single estimate of weights** (vulnerable to overfitting)

  => look at the **distribution over weights** given data i.e. **posterior** $p(\theta|\mathcal{D})$

- For input **x**, estimate the **predictive distribution** for output **y** by **averaging over** all possible **models** weighted by the **posterior** dist. over the weights:

$$\underbrace{p(\mathbf{y}|\mathbf{x},\mathcal{D})}_{\text{Predictive distribution}} = \int \underbrace{p(\mathbf{y}|\mathbf{x},\theta,\mathcal{D})}_{\text{Likelihood model}}\underbrace{p(\theta|\mathcal{D})}_{\text{Posterior over weights}}d\theta$$

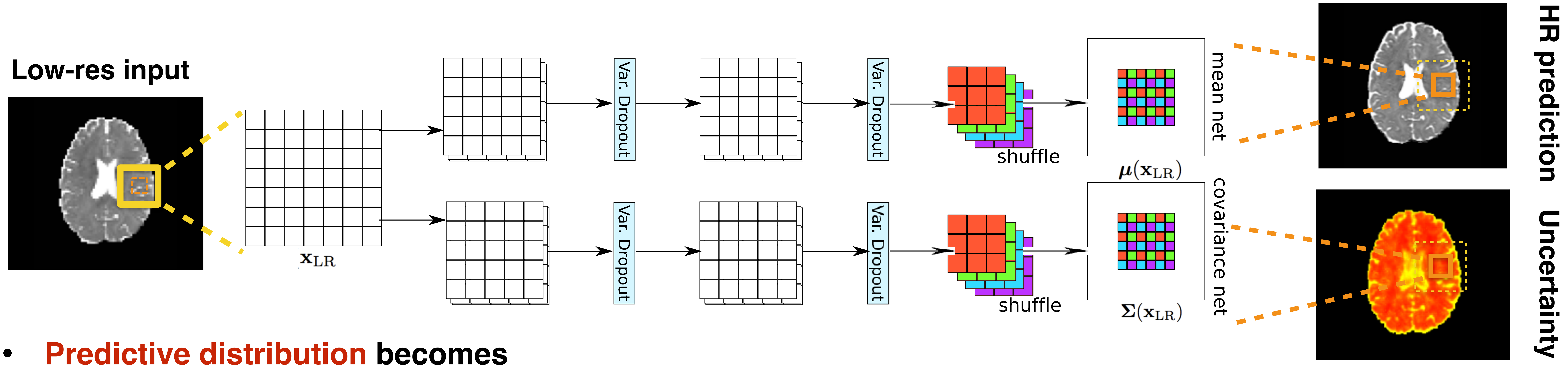
- **But** …. **posterior** $p(\theta|\mathcal{D})$ is intractable

  => approximate with a Gaussian dist. $q_\phi(\theta)$ using **Variational Dropout** **[Kingma et al. NIPS'15]**

- **Why Variational Dropout?**

  => **Dropout probabilities are learned** during training: **no grid search** is required.

Combine intrinsic uncertainty and parameter uncertainty



- **Predictive distribution becomes**

$$p(\mathbf{y}|\mathbf{x}, \mathcal{D}) \approx \int \underbrace{\mathcal{N}(\mathbf{y}; \mu_{\theta_1}(\mathbf{x}), \Sigma_{\theta_2}(\mathbf{x}))}_{\substack{\text{Likelihood} \\ \text{hetero. noise model} \\ \text{intrinsic uncertainty}}} \cdot \underbrace{q_\phi(\theta)}_{\substack{\text{Posterior} \\ \text{var. dropout} \\ \text{parameter uncertainty}}} d\theta$$

- **MC dropout test time:** run multiple forward passes and collect many samples $\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, ..., \mathbf{y}^{(T)}\}$
- Estimate the **predictive mean** and **predictive uncertainty** (variance).

Method Evaluation Strategy

We will show the proposed deep learning methods:

(I) are more **accurate and faster** on 3D super-resolution of DTIs.

(II) **benefit tractography** through super-resolution of MAP-MRI.

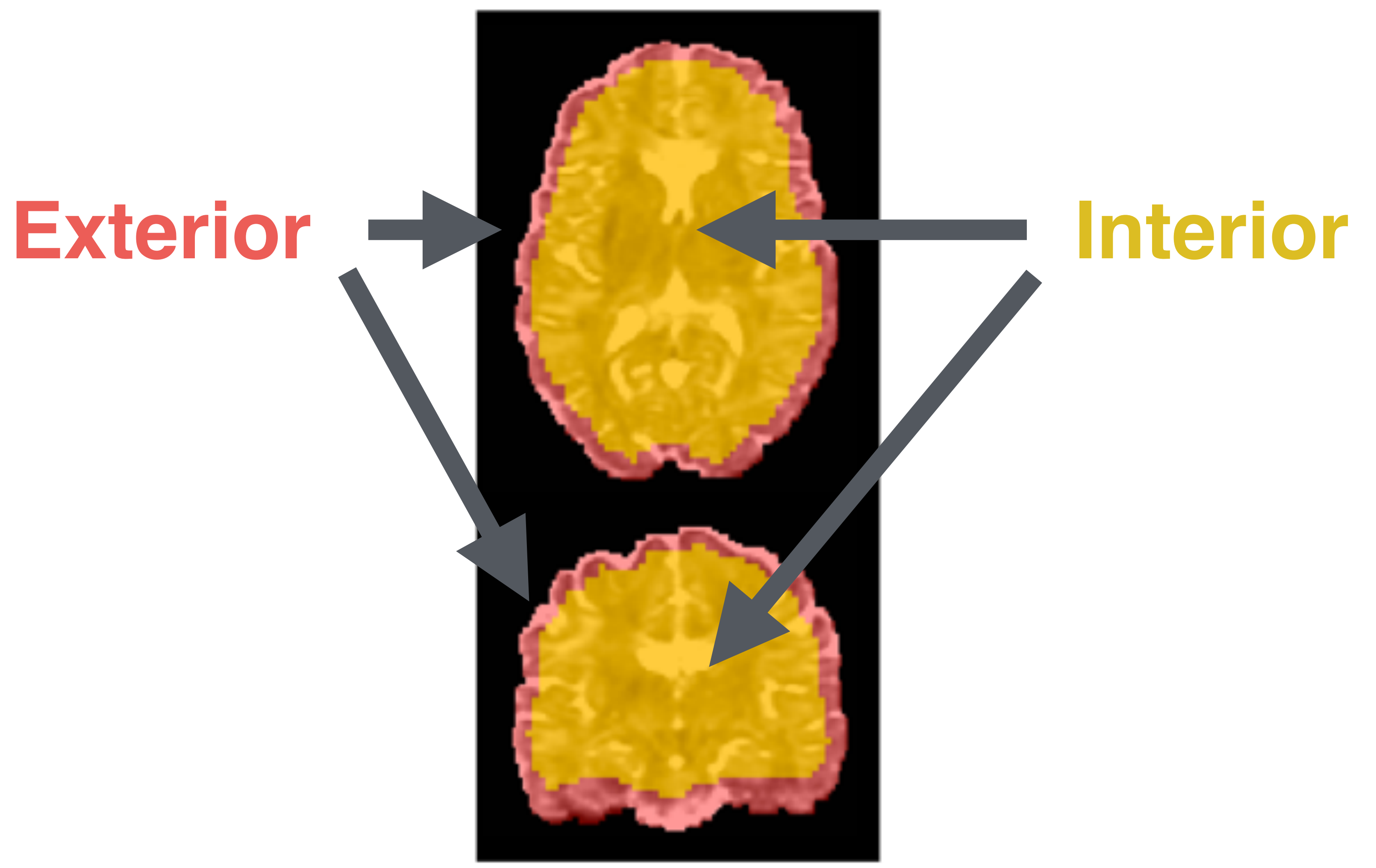
(III) produce a useful estimate of **predictive uncertainty.**

Experiment (I): x2 DTI Super-resolution (2.5 mm => 1.25 mm)
Comparison on HCP and Lifespan dataset

- **Trained** on 8 randomly selected subjects from HCP dataset (age 22 - 36)
  (low-res = 2.5 mm and high-res =1.25 mm isotropic voxels)

- **Evaluated** performance on two datasets
  - (a) **(within train dist.):** 8 unseen subjects from the same HCP cohort.
  - (b) **(outside train dist.):** 10 subjects from Lifespan dataset (older age 45 - 75, different protocol)

- **Computed errors:** Root-Mean-Squared-Error (RMSE) on the **interior** and **exterior** regions separately.

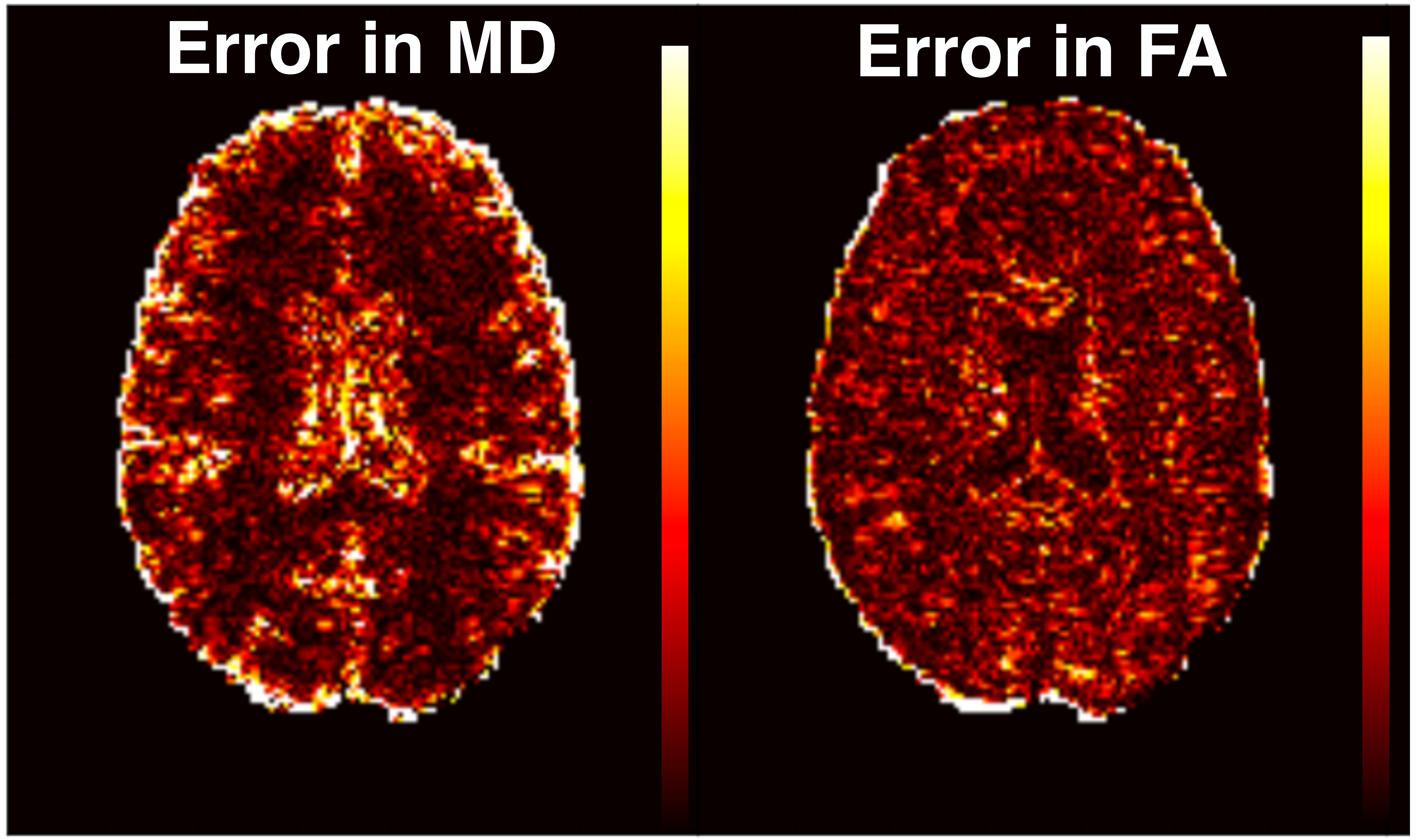


Experiment (I): x2 DTI Super-resolution (2.5 mm => 1.25 mm)
Comparison on HCP and Lifespan dataset

RMSE (mm^2s^{-1})

Models	HCP (interior)	HCP (exterior)	Life (interior)	Life (exterior)
Cubic interpolation	10.069± n/a	31.738± n/a	32.483± n/a	49.066± n/a
BIQT-Random-Forests (published best method)	6.972 ± 0.069	23.110 ± 0.362	9.926 ± 0.055	25.208 ± 0.290
3D-ESPCN(baseline network)	6.378 ± 0.015	13.909 ± 0.071	8.998 ± 0.021	16.779 ± 0.109

BIQT-Random Forests

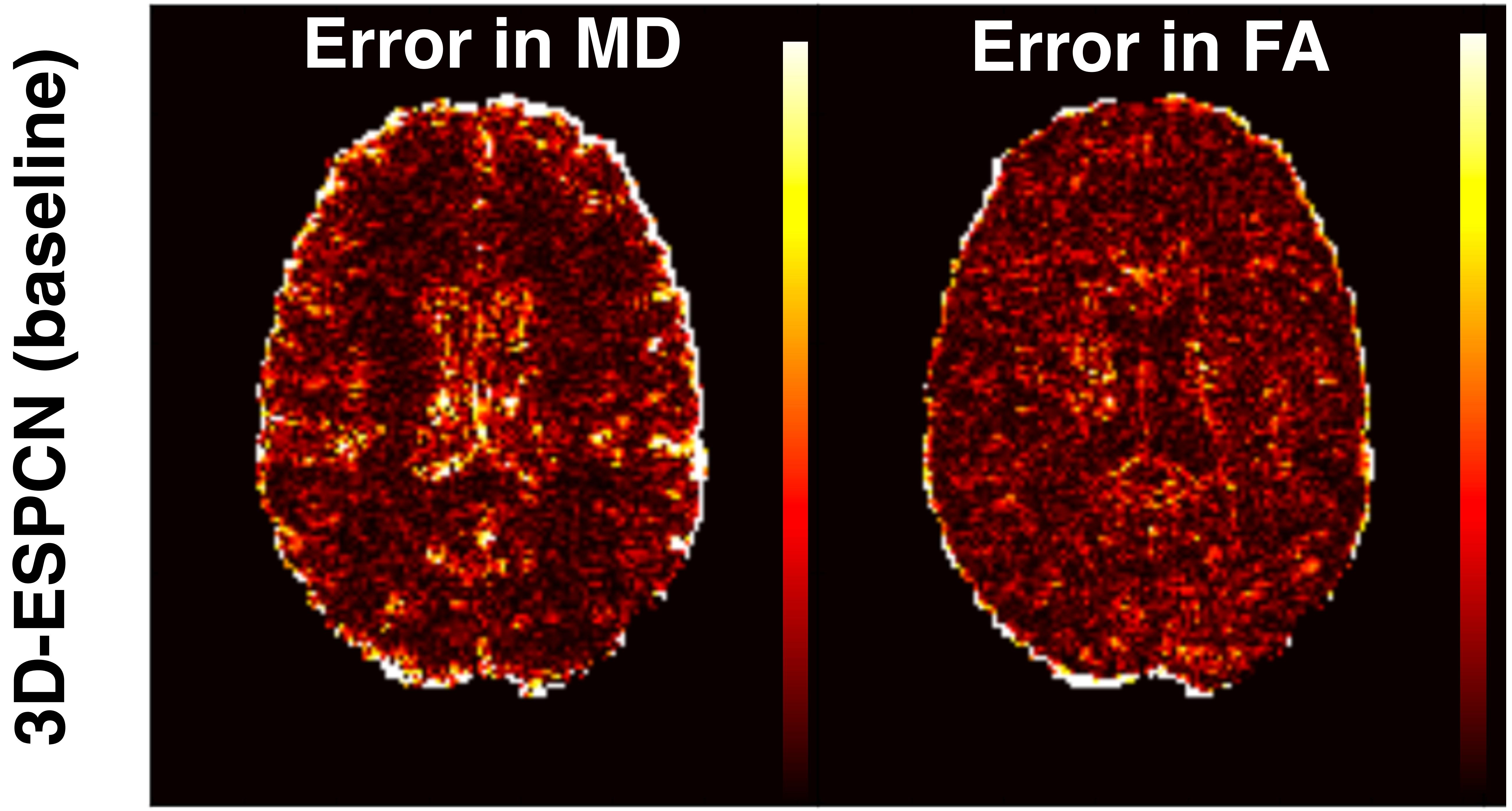


- **HCP**: 3D-ESPCN: 8.5%👇(interior), 39.8%👇(exterior) reduction in RMSE from BIQT-RF, $p < 0.001$
  Lifespan: 3D-ESPCN: 9.3%👇 (interior), 33.4%👇(exterior), $p < 0.001$
- **Very fast**: 1s on a GPU and 10s on a CPU while BIQT-RF takes 10 mins.

Experiment (I): x2 DTI Super-resolution (2.5 mm => 1.25 mm)
Comparison on HCP and Lifespan dataset

RMSE ($mm^2 s^{-1}$)

Models	HCP (interior)	HCP (exterior)	Life (interior)	Life (exterior)
Cubic interpolation	10.069± n/a	31.738± n/a	32.483± n/a	49.066± n/a
BIQT-Random-Forests (published best method)	6.972 ± 0.069	23.110 ± 0.362	9.926 ± 0.055	25.208 ± 0.290
3D-ESPCN(baseline network)	6.378 ± 0.015	13.909 ± 0.071	8.998 ± 0.021	16.779 ± 0.109



- **HCP**: 3D-ESPCN: 8.5%👇(interior), 39.8%👇(exterior) reduction in RMSE from BIQT-RF, $p < 0.001$
  Lifespan: 3D-ESPCN: 9.3%👇 (interior), 33.4%👇(exterior), $p < 0.001$
- **Very fast**: 1s on a GPU and 10s on a CPU while BIQT-RF takes 10 mins.

Experiment (I): x2 DTI Super-resolution (2.5 mm => 1.25 mm)
Comparison on HCP and Lifespan dataset

RMSE (mm^2s^{-1})

CNN based models

Models	HCP (interior)	HCP (exterior)	Life (interior)	Life (exterior)
Cubic interpolation	10.069± n/a	31.738± n/a	32.483± n/a	49.066± n/a
BIQT-Random-Forests (published best method)	6.972 ± 0.069	23.110 ± 0.362	9.926 ± 0.055	25.208 ± 0.290
3D-ESPCN(baseline network)	6.378 ± 0.015	13.909 ± 0.071	8.998 ± 0.021	16.779 ± 0.109
Hetero-Noise-CNN	6.294 ± 0.029	15.569 ± 0.273	8.985 ± 0.051	17.716 ± 0.277
Variational-Dropout (I)-CNN	6.354 ± 0.015	$\mathbf{13.824 \pm 0.031}$	8.973 ± 0.024	$\mathbf{16.633 \pm 0.053}$
Variational-Dropout (II)-CNN	6.356 ± 0.008	13.846 ± 0.017	8.982 ± 0.024	16.738 ± 0.073
Hetero-Noise-CNN+Variational-Dropout (I)	6.291 ± 0.012	13.906 ± 0.048	$\mathbf{8.944 \pm 0.044}$	16.761 ± 0.047
Hetero-Noise-CNN+Variational-Dropout (II)	$\mathbf{6.287 \pm 0.029}$	13.927 ± 0.093	8.955 ± 0.029	16.844 ± 0.109

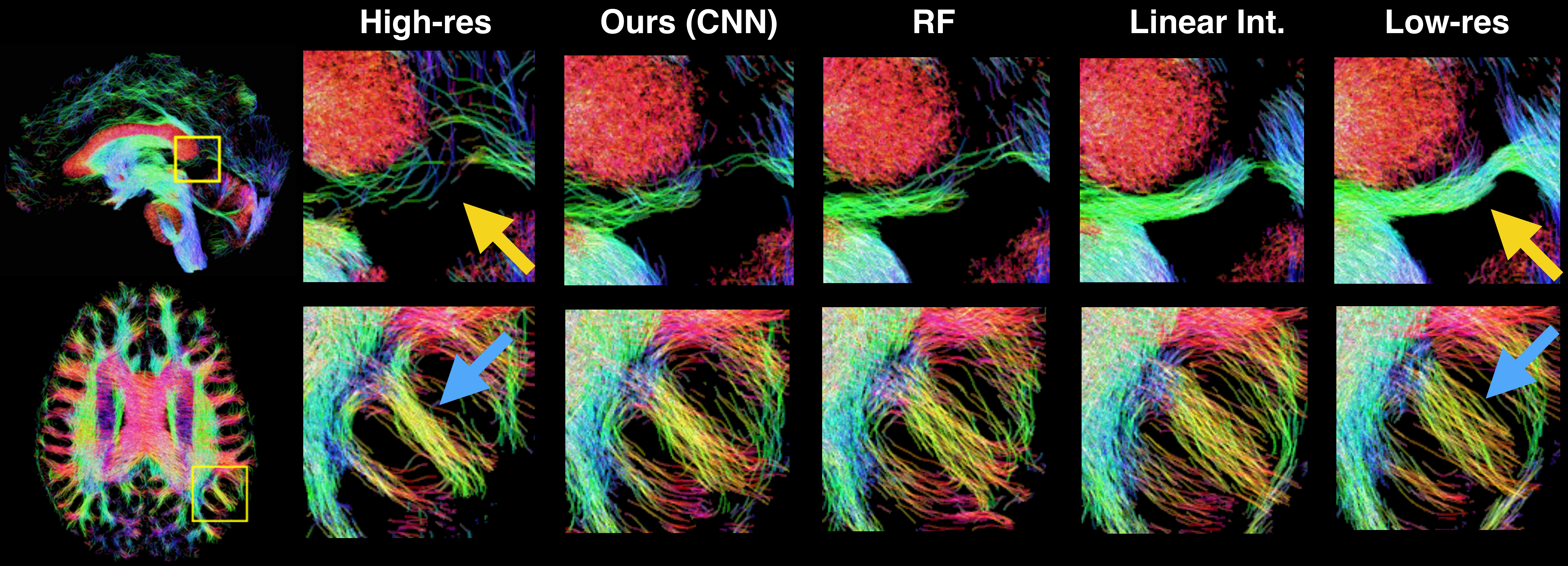
best

2nd best

- **TOP2 models:** Hetero-Noise + Variational-Dropout (interior) & Variational-Dropout only (exterior) (better than the baseline with $p<0.001$)

Experiment (II): Benefits in Tractography

Separate high-res and low-res acquisitions, "Prisma" dataset, [Alexander et al., NIMG'17]



- (yellow arrows): avoids the false positive tract under the corpus callosum
- (blue arrows): shaper recovery of small gyral white matter pathways

Experiment (III): Predictive Uncertainty

Comparison on a test HCP subject

- Mean and std estimated from 200 samples of predicted high-res DTIs with Hetero+Var.(I) model

Experiment (III): Predictive Uncertainty
Testing on a clinical image of a brain tumour patient

Clinical image | After Super-res | Uncertainty | high

low

- Used the best model: Hetero + Var. (II)
- Highlights pathology with high uncertainty

Take Home Messages

- A minimal CNN model achieves state-of-the-art performance and speed in super-resolution of dMRI, with tangible benefits in tractography.

- Modelling intrinsic and parameter uncertainty improves accuracy.

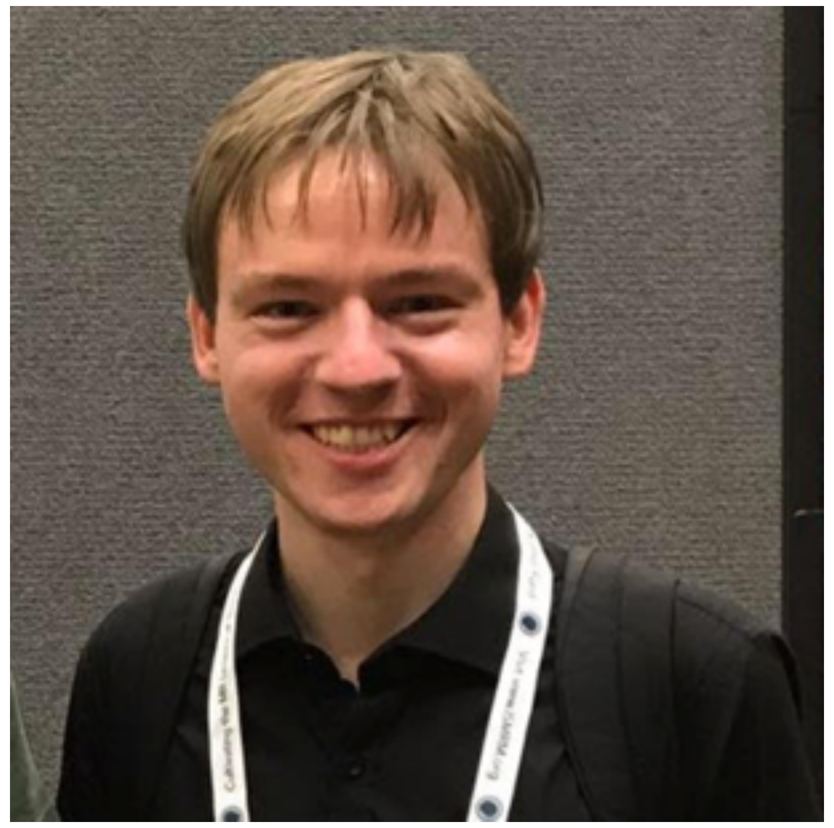
- Predictive uncertainty can be potentially used as a safeguard against failures in predictions.

- Applicable to many other image analysis problems

Acknowledgements

Daniel Worrall

Aurobrata Ghosh



Enrico Kaden

Stamatios Sotiropoulos

Antonio Criminisi

Daniel C. Alexander

Microsoft Research Cambridge

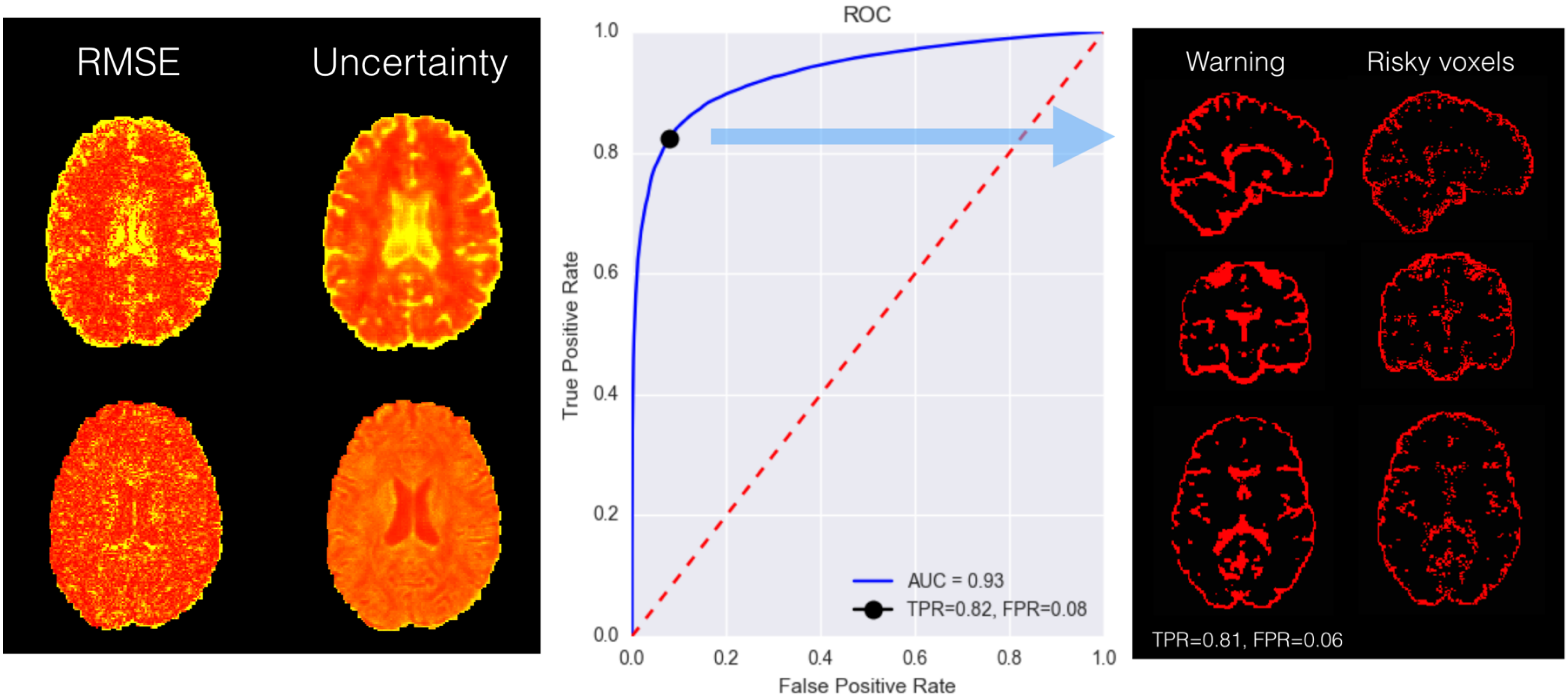
cmic Centre for Medical Image Computing

EPSRC Engineering and Physical Sciences Research Council

NHS National Institute for Health Research

Come and talk to us!
poster #70
today @10:30 am

Appendix: risk assessment with predictive uncertainty



- Can discriminate risky voxels with 94% accuracy

Appendix: decompose predictive uncertainty into sources

- **Predictive uncertainty** arises from the combination of two sources: **intrinsic** and **parameter** uncertainty.
- **Decompose** the predictive uncertainty into two sources:

$$\begin{aligned}\mathbb{V}_{p(\mathbf{y}|\mathbf{x},\mathcal{D})}[\mathbf{y}] &= \mathbb{E}_{p(\theta|\mathcal{D})}\left[\mathbb{V}_{p(\mathbf{y}|\mathbf{x},\mathcal{D})}[\mathbf{y}] - \mathbb{V}_{p(\mathbf{y}|\theta,\mathbf{x},\mathcal{D})}[\mathbf{y}|\theta]\right] + \mathbb{E}_{p(\theta|\mathcal{D})}\left[\mathbb{V}_{p(\mathbf{y}|\theta,\mathbf{x},\mathcal{D})}[\mathbf{y}|\theta]\right] \\ &= \underbrace{\mathbb{V}_{p(\theta|\mathcal{D})}[\mathbb{E}_{p(\mathbf{y}|\theta,\mathbf{x},\mathcal{D})}[\mathbf{y}|\theta]]}_{\text{propagated parameter uncertainty}} + \underbrace{\mathbb{E}_{p(\theta|\mathcal{D})}[\mathbb{V}_{p(\mathbf{y}|\theta,\mathbf{x},\mathcal{D})}[\mathbf{y}|\theta]]}_{\text{propagated intrinsic uncertainty}} \\ &= \Delta_m(\mathbf{y}) + \Delta_i(\mathbf{y})\end{aligned}$$

- **Simplify** this for the heteroscedastic + var. dropout model that:

$$\begin{aligned}\Delta_m(\mathbf{y}) &= \mathbb{V}_{p(\theta_1|\mathcal{D})}[\mu_{\theta_1}(\mathbf{x})] \\ \Delta_i(\mathbf{y}) &= \mathbb{E}_{p(\theta_2|\mathcal{D})}[\Sigma_{\theta_2}(\mathbf{x})]\end{aligned}$$

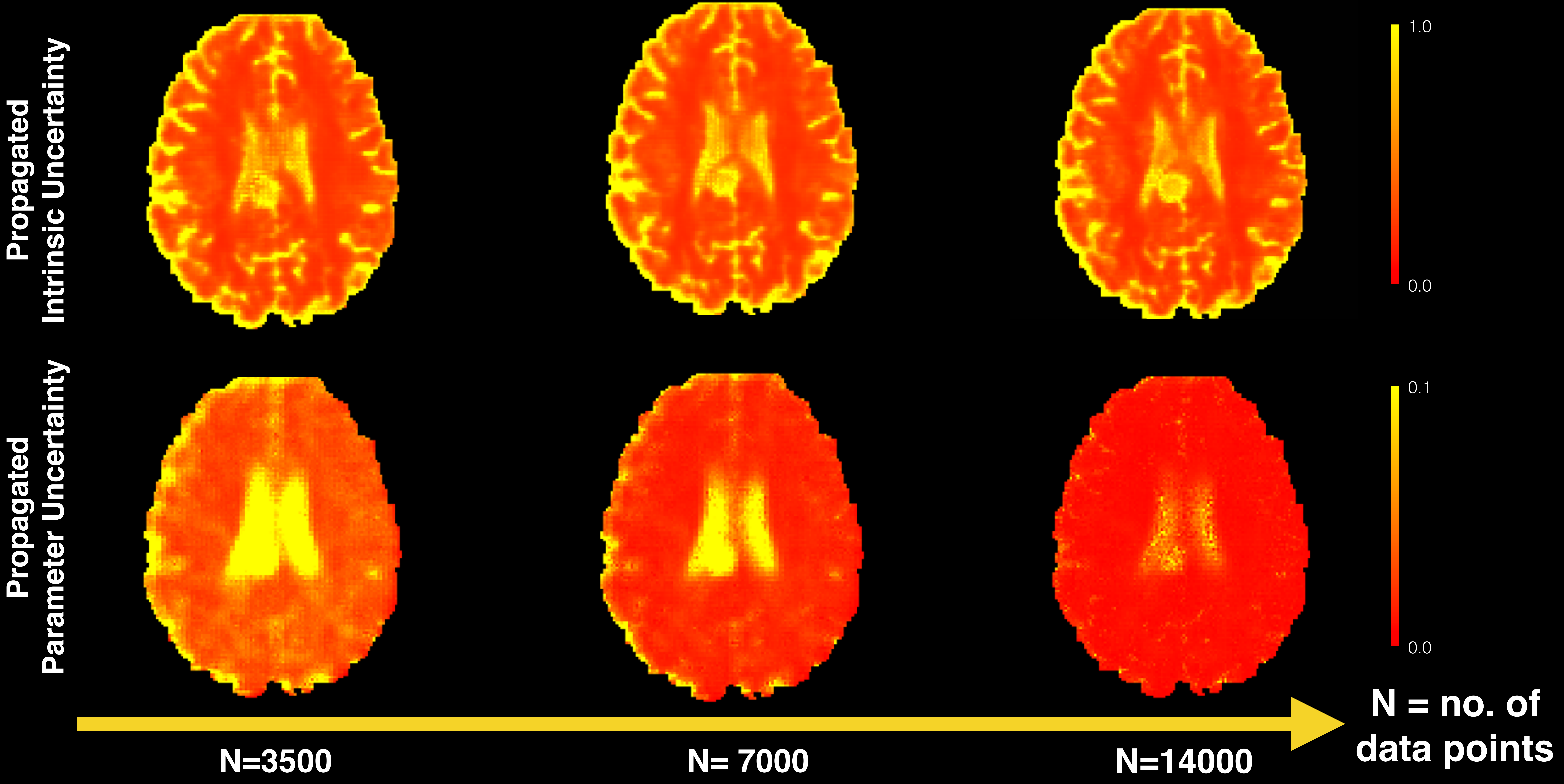
- **Unbiased MC estimators:**

$$\widehat{\Delta}_m(\mathbf{y}) = \frac{1}{T}\sum_{t=1}^{T}\mu_{\theta_1^t}(\mathbf{x})\mu_{\theta_1^t}(\mathbf{x})^T - \hat{\mu}_{\mathbf{y}|\mathbf{x}}\hat{\mu}_{\mathbf{y}|\mathbf{x}}^T$$

$$\widehat{\Delta}_i(\mathbf{y}) = \frac{1}{T}\sum_{t=1}^{T}\Sigma_{\theta_2^t}(\mathbf{x})$$

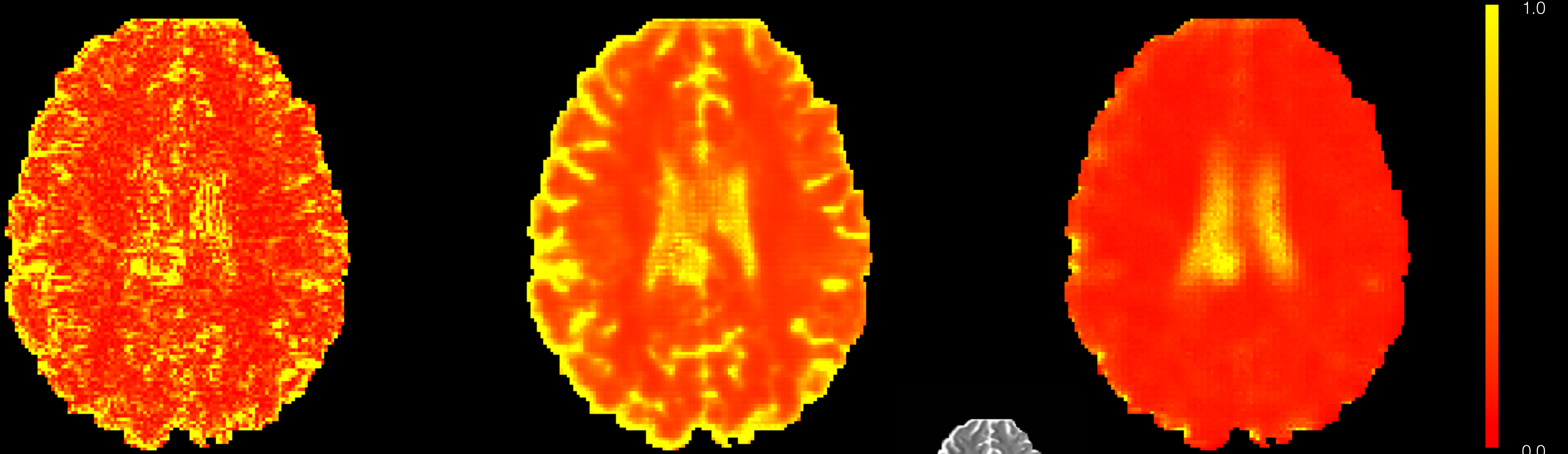
Appendix: Decomposition of predictive uncertainty

Training data size vs uncertainty components

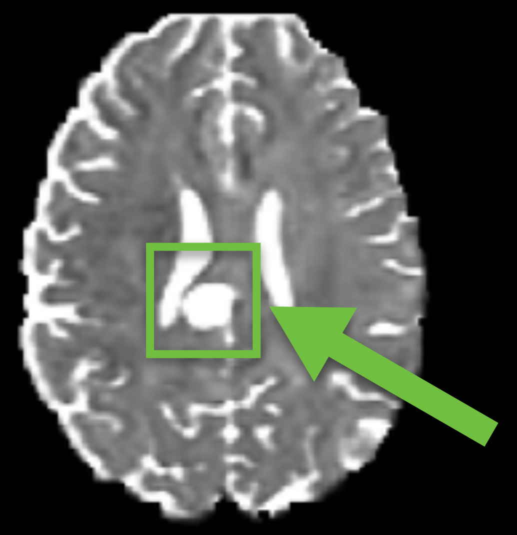


Appendix: Decomposition of predictive uncertainty

Error (RMSE) | Propagated Intrinsic Uncertainty | Propagated Parameter Uncertainty

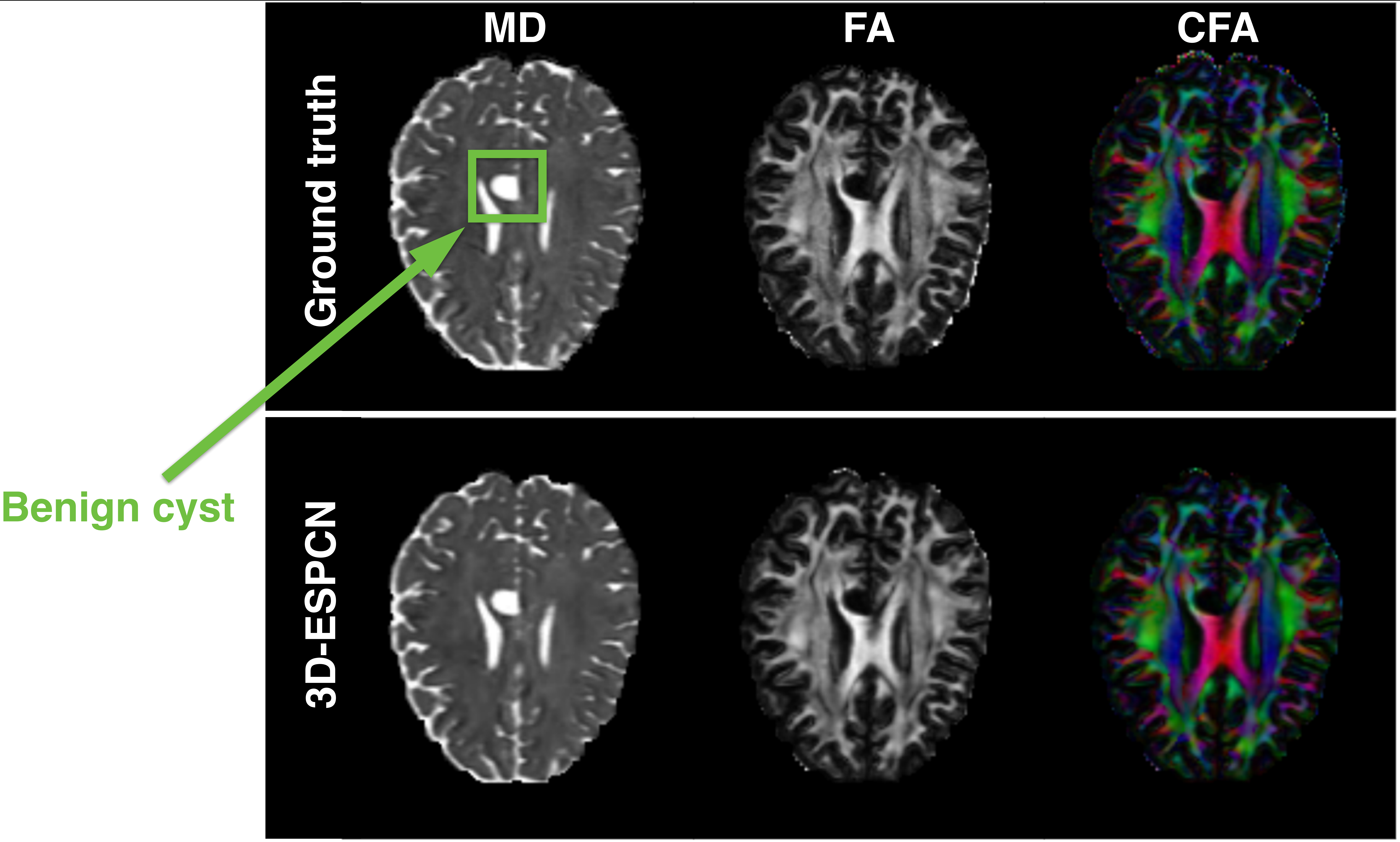


- Small training set (~ 3000 patch pairs)



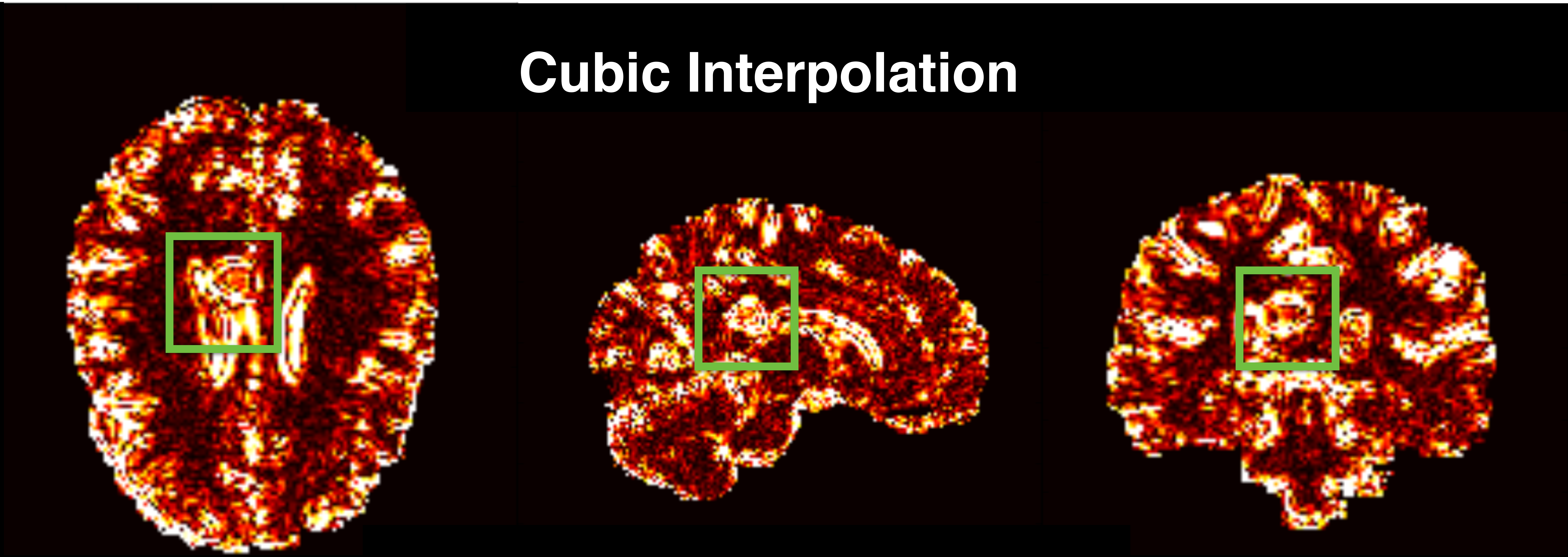
Benign cyst

Appendix: performance on abnormality (1/2)



Appendix: performance on abnormality (2/2)

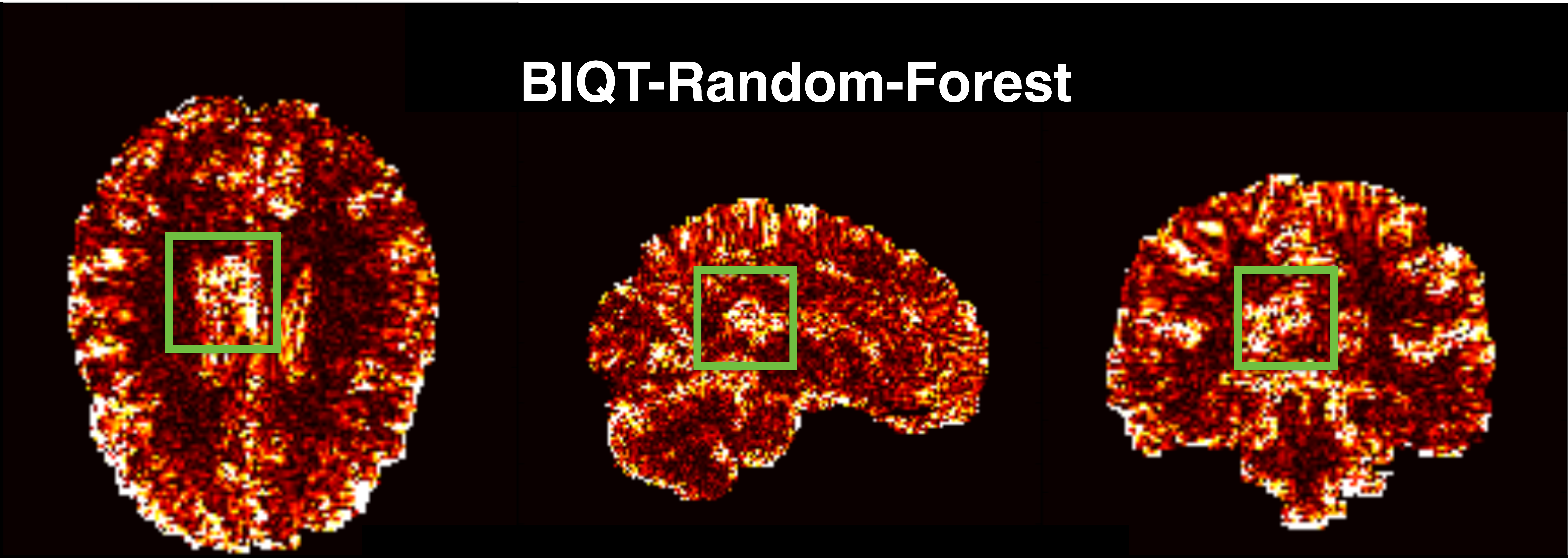
Showing RMSE in Mean Diffusivity



Appendix: performance on abnormality (2/2)

Showing RMSE in Mean Diffusivity

BIQT-Random-Forest



Appendix: performance on abnormality (2/2)

Showing RMSE in Mean Diffusivity