In this supplementary material, we derive the estimation of synthetic CT patches $\mathbf{u}_i$. The Expectation-Maximization framework takes the perspective of an incomplete versus complete data problem to compute the best matching patch. We define z_{it} to be the indicator function that $\mathbf{p}_i$ comes from a GMM of the $t = \{j,k\}^{\text{th}}$ patch pair, under the constraints that $\sum_{t\in\Psi} z_{it} = 1\ \forall i$, $z_{it} \in \{0,1\}$. If we know the values of the indicator functions (the complete data), we would know what is the matching patch and we would also be able to estimate the unknown covariance matrices Σ_t and the unknown weighting coefficients α_{it}. The probability of observing $\mathbf{p}_i$ can be written as,

$$P(\mathbf{p}_i|z_{it}=1;\Sigma_t,\alpha_{it}) = \frac{1}{\sqrt{2\pi}|\Sigma_t|}\exp\left\{-\frac{1}{2}\mathbf{h}_{it}^T\Sigma_t^{-1}\mathbf{h}_{it}\right\}, \tag{1}$$

where $\mathbf{h}_{it} = \mathbf{p}_i - \alpha_{it}\mathbf{q}_j - (1-\alpha_{it})\mathbf{q}_k, t \equiv \{j,k\}$. Although we are not provided the indicator functions, the EM framework allows us to estimate the parameters by computing expectations of the indicator functions. We assume that the mixing coefficients are equal for each component of the mixture model.

We have experimentally found that any arbitrary positive definite Σ_t allows for too many degrees of freedom and is not robust to estimate. To simplify the problem, we assume it to be separable and diagonal and is given by

$$\Sigma_t = \begin{bmatrix} \sigma_{1t}^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_{2t}^2\mathbf{I} \end{bmatrix},$$

indicating that the variations of each voxel in a patch are the same around the means, although individual voxels can be of different tissues. Thus the joint probability becomes

$$P(\mathbf{P},\mathbf{Z};\mathbf{\Theta}) = D\prod_{t\in\Psi}\prod_{i=1}^{N}\left[\frac{1}{\sigma_{1t}\sigma_{2t}}\exp\left\{-\frac{||\mathbf{f}_{it}||^2}{2\sigma_{1t}^2}\right\}\exp\left\{-\frac{||\mathbf{g}_{it}||^2}{2\sigma_{2t}^2}\right\}\right]^{z_{it}},$$
$$\mathbf{f}_{it} = \mathbf{x}_i - \alpha_{it}\mathbf{y}_j - (1-\alpha_{it})\mathbf{y}_k, \quad \mathbf{g}_{it} = \mathbf{u}_i - \alpha_{it}\mathbf{v}_j - (1-\alpha_{it})\mathbf{v}_k. \tag{2}$$

The set of unknown parameters are $\mathbf{\Theta} = \{\sigma_{1t},\sigma_{2t},\alpha_{it}\}$, and $\mathbf{Z} = \{z_{it}\}, i = 1,\ldots,N, t\in\Psi$.

The maximum likelihood estimators of $\mathbf{\Theta}$ are found by maximizing Eqn. 2 using EM. The **E**-step requires the computation of $E(z_{it}|\mathbf{P},\mathbf{\Theta}^{(m)}) = P(z_{it}|\mathbf{P},\mathbf{\Theta}^{(m)})$. Given that z_{it} is an indicator function, it can be shown that $E(z_{it}|\mathbf{P},\mathbf{\Theta}^{(m)}) = w_{it}^{(m)}$, where

$$w_{it}^{(m+1)} = \frac{\frac{1}{\sigma_{1t}^{(m)}\sigma_{2t}^{(m)}}\exp\left\{-\frac{||\mathbf{f}_{it}^{(m)}||^2}{2\sigma_{1t}^{(m)^2}}\right\}\exp\left\{-\frac{||\mathbf{g}_{it}^{(m)}||^2}{2\sigma_{2t}^{(m)^2}}\right\}}{\sum_{\ell\in\Psi}\frac{1}{\sigma_{1\ell}^{(m)}\sigma_{2\ell}^{(m)}}\exp\left\{-\frac{||\mathbf{f}_{i\ell}^{(m)}||^2}{2\sigma_{1\ell}^{(m)^2}}\right\}\exp\left\{-\frac{||\mathbf{g}_{i\ell}^{(m)}||^2}{2\sigma_{2\ell}^{(m)^2}}\right\}}, \tag{3}$$

$w_{it}^{(m)}$ being the posterior probability of $\mathbf{p}_i$ originating from the Gaussian distribution of the t^{th} reference patches $\mathbf{q}_j$ and $\mathbf{q}_k$. $\mathbf{f}_{it}^{(m)}$ and $\mathbf{g}_{it}^{(m)}$ are the expressions defined in Eqn. 2 but with $\alpha_{it}^{(m)}$. $\mathbf{f}_{i\ell}^{(m)}$ and $\mathbf{g}_{i\ell}^{(m)}$ denote the corresponding values with reference patches belonging to the ℓ^{th} pair, $\ell \in \Psi$, with $\alpha_{i\ell}^{(m)}$. The synthetic patches are obtained by the following expectation,

$$E(\mathbf{u}_i|\mathbf{\Theta}^{(m)}) = \sum_{t \in \Psi} w_{it}^{(m)} \left(\alpha_{it}^{(m)} \mathbf{v}_j + (1 - \alpha_{it}^{(m)}) \mathbf{v}_k \right).$$

At each iteration, we replace the value of $\mathbf{u}_i$ with its expectation. The **M**-step involves the maximization of the log of the expectation with respect to the parameters given the current $w_{it}^{(m)}$. The update equations are given by,

$$\sigma_{1t}^{(m+1)^2} = \frac{\sum_{i=1}^{N} w_{it}^{(m)} ||\mathbf{x}_i - \alpha_{it}^{(m)} \mathbf{y}_j - (1 - \alpha_{it}^{(m)}) \mathbf{y}_k||^2}{\sum_{i=1}^{N} w_{it}^{(m)}}, \tag{4}$$

$$\sigma_{2t}^{(m+1)^2} = \frac{\sum_{i=1}^{N} w_{it}^{(m)} ||\mathbf{u}_i - \alpha_{it}^{(m)} \mathbf{v}_j - (1 - \alpha_{it}^{(m)}) \mathbf{v}_k||^2}{\sum_{i=1}^{N} w_{it}^{(m)}}, \tag{5}$$

$$\begin{aligned}
&\alpha_{it}^{(m)} : F\left(\alpha_{it}^{(m)}\right) = 0, \text{where } F(x) = Ax^2(1-x) - Bx(1-x) + 2x - 1, \\
&A = \frac{||\mathbf{y}_k - \mathbf{y}_j||^2}{\sigma_{1t}^{(m)^2}} + \frac{||\mathbf{v}_k - \mathbf{v}_j||^2}{\sigma_{2t}^{(m)^2}}, \\
&B = \frac{(\mathbf{y}_k - \mathbf{x}_i)^T (\mathbf{y}_k - \mathbf{y}_j)}{\sigma_{1t}^{(m)^2}} + \frac{(\mathbf{v}_k - \mathbf{u}_i)^T (\mathbf{v}_k - \mathbf{v}_j)}{\sigma_{2t}^{(m)^2}}.
\end{aligned} \tag{6}$$

It should be noted that $F(0) = -1, F(1) = 1, \forall\ A, B$, thus there is always a feasible $\alpha_{it}^{(m)} \in (0,1)$. The EM algorithm is said to converge at iteration m, if $||\mathbf{u}_i^{(m+1)} - \mathbf{u}_i^{(m)}|| < \delta$ for some small δ. Once the EM algorithm has converged, the expectation of the final $\mathbf{u}_i$ is considered the synthetic CT patch, and the center voxel of $\mathbf{u}_i$ is used as the CT replacement of the i^{th} voxel.

The model assumes all possible pairs $\binom{M}{2}$ of reference patches. Thus the complexity of the model is $\mathcal{O}(NM^2)$. For a typical 1mm^3 UTE scan, $M, N \sim 10^7$. Thus it is almost infeasible to solve with all possible reference patches. However, the Gaussian model is valid for those reference and subject patches that are close in intensity. Using a non-local type of criteria, for every subject patch $\mathbf{x}_i$, we start with a feasible set of L reference patches such that they are the L nearest neighbors of $\mathbf{x}_i$. Thus the i^{th} subject patch follows an $\binom{L}{2}$-class GMM and the complexity becomes $\mathcal{O}(NL^2)$. In all our experiments, we choose $3 \times 3 \times 3$ patches ($d = 27$), $L = 20$, and $\delta = 0.001 \max(a_3)$.