# Alternative Learning Paradigms for Image Quality Transfer

Ahmed Karam Eldaly https://orcid.org/0000-0002-2266-6992 A.Karam@ucl.ac.uk
Centre for Medical Image Computing,
Department of Computer Science, University College London, UK
Department of Computer Science, University of Exeter, UK

Matteo Figini https://orcid.org/0000-0002-8238-2262 m.figini@ucl.ac.uk
Centre for Medical Image Computing,
Department of Computer Science, University College London, UK

Daniel C. Alexander https://orcid.org/0000-0003-2439-350X D.Alexander@ucl.ac.uk
Centre for Medical Image Computing, Department of Computer Science, University College London, UK

## Abstract

Image Quality Transfer (IQT) aims to enhance the contrast and resolution of low-quality medical images, e.g. obtained from low-power devices, with rich information learned from higher quality images. In contrast to existing IQT methods in the literature which adopt supervised learning frameworks, in this work, we propose two novel formulations of the IQT problem. The first approach uses an unsupervised learning framework, whereas the second is a combination of both supervised and unsupervised learning. The unsupervised learning approach considers a sparse representation (SRep) and dictionary learning model, which we call IQT-SRep, whereas the combination of supervised and unsupervised learning approach is based on deep dictionary learning (DDL), which we call IQT-DDL. The IQT-SRep approach trains two dictionaries using a sparse representation model using pairs of low- and high-quality volumes. Subsequently, the sparse representation of a low-quality block, in terms of the low-quality dictionary, can be directly used to recover the corresponding high-quality block using the high-quality dictionary. On the other hand, the IQT-DDL approach explicitly learns a high-resolution dictionary to upscale the input volume, while the entire network, including high dictionary generator, is simultaneously optimised to take full advantage of deep learning methods. The two models are evaluated using a low-field magnetic resonance imaging (MRI) application aiming to recover high-quality images akin to those obtained from high-field scanners. Experiments comparing the proposed approaches against state-of-the-art supervised deep learning IQT method (IQT-DL) identify that the two novel formulations of the IQT problem can avoid bias associated with supervised methods when tested using out-of-distribution data that differs from the distribution of the data the model was trained on. This highlights the potential benefit of these novel paradigms for IQT.

**Keywords:** Image Quality Transfer, Supervised Learning, Unsupervised Learning, Sparse Representation, Dictionary Learning, Deep Dictionary Learning, Deep Learning, Out-of-Distribution, In-distribution

## 1. Introduction

Image Quality Transfer (IQT) Alexander et al. (2014, 2017); Lin et al. (2019); Tanno et al. (2021); Lin et al. (2021, 2022); Kim et al. (2023) is a machine learning technique that is used to enhance the resolution and contrast of low-quality clinical data using rich information in high-quality images. For example given an image from a standard hospital scanner or rapid acquisition protocol, we might estimate the image we would have got from the same subject using a high-power experimental scanner available only in specialist research centres or a richer acquisition protocols too lengthy to run on every patient. IQT is a vital component of efforts to democratise the capabilities of high power rare experimental systems broadening the accessibility e.g. to lower and middle income countries Anazodo et al. (2022). This technique learns mappings from low-quality (e.g. clinical) to high-quality (e.g.experimental) images exploiting the similarity of image structure across subjects, regions, modalities, and scales. The mapping may then operate directly on low-quality images to estimate the corresponding high-quality images. Early work Alexander et al. (2017); Blumberg et al. (2018); Tanno et al. (2021) focused on diffusion MRI and showed remarkable ability to enhance both contrast and resolution and enabled tractography to recover small pathways impossible to reconstruct at the acquired resolution. Recent work Lin et al. (2021) extends the idea to standard structural MRI, particularly targeting application to low-field MRI systems. IQT technique Alexander et al. (2017) differs from super-resolution in computer vision Lau et al. (2023); Zhou et al. (2020, 2021); Li et al. (2024) in several key aspects. In general super-resolution aim to up-sample an image, whereas IQT aims to transfer the quality of information from an image to the other. This means that IQT is not limited to increasing the spatial resolution of images. While super-resolution techniques primarily focus on enhancing the spatial resolution, IQT also aims to improve the image contrast. This dual enhancement is crucial for medical imaging applications where both resolution and contrast are necessary for accurate diagnosis and analysis. Moreover, super-resolution techniques are generally used to upsample images, making them appear sharper and more detailed. In contrast, IQT is specifically designed to transfer the quality from high-quality images to low-quality images. This is particularly beneficial in medical imaging, where high-quality images from advanced scanners are used to enhance the quality of images obtained from lower-power or less advanced scanners. Lastly, IQT differs from modality transfer methods, which maps one modality to another to obtain multi-modality information Iglesias et al. (2021, 2023, 2022), whereas IQT's primary goal is to enhance the existing image quality, specifically improving resolution and contrast rather than the developing new content. By highlighting these differences, we aim to clearly delineate the unique characteristics and advantages of the IQT task.

Machine learning models are often trained on a specific data distribution, but may encounter unseen data from different distributions in real-world scenarios. This poses a critical challenge for the security and reliability of machine learning systems, especially in some error-sensitive applications, such as medical diagnosis including the application investigated in this work. One of its powerful capabilities lies in the promising generalisation ability from training data to unseen in-distribution (InD) data. However, the finite training data cannot guarantee the completeness of data distribution, so it is inevitable to encounter out-of-distribution (OOD) data. Machine learning models can be broadly categorised into supervised, unsupervised and self-supervised learning models. In supervised learning, the model is trained by paring inputs with their expected outputs. However, this is far from being practical, since the full data distribution cannot be represented in the training data set. To circumvent this difficulty, unsupervised and self-supervised learning methods can be used.

All IQT models proposed in the literature use supervised learning frameworks to learn a regression between matched patches in low- and high-quality images. In particular deep learning frameworks substantially outperform the original random-forest implementation in terms of global error metrics for enhancement of both diffusion-tensor MRI and low-field structural MRI Alexander et al. (2014, 2017); Lin et al. (2019); Tanno et al. (2021); Lin et al. (2021, 2022). However, interpretation of images enhanced via such regression models needs caution. First, regression models in general can lead to bias that depends on the training data distribution Obermeyer et al. (2019). In particular, inputs (here patches) that are rare in the training data are often skewed towards outputs more common in training data; and degenerate regions of the input-space where the mapping is ambiguous are often mapped to a consistent mean giving a false impression of consistent and confident output. Moreover, the performance of deep-learning based methods can degrade even more with OOD data. These effects have been well documented in other image-related regression applications recently, such as parameter mapping Gyori et al. (2022). So far, they have not been considered in IQT and image enhancement although similar effects are likely to arise. Additional problems, particularly in deep learning, can arise from over-fitting and under-fitting which can further add to bias in estimates particularly for examples that are over/under-represented in the training data. Moreover, state-of-the-art IQT models, specifically deep neural networks, are generally designed for a static and closed world Krizhevsky et al. (2017); He et al. (2015). The models are trained under the assumption that the input distribution at test time will be the same as the training distribution. In real world MRI data, however, deep-learning-based techniques effectiveness diminishes when applied to images that differ significantly from the training data set Gu et al. (2019). Although various approaches have been developed to tackle this issue, such as training networks to handle multiple types of degradation Soh et al. (2020); Xu et al. (2020); Zhang et al. (2018a); Zhou and Susstrunk (2019) and making models less sensitive to degradation through iterative optimisations Shocher et al. (2018); Gu et al. (2019), it is also crucial to enhance the robustness of the network structure.

Sparse representation (SRep) using dictionary learning is an unsupervised learning framework that assumes a given signal is sparse in some domain (Wavelets, Fourier, discrete cosine transform, etc.). SRep has proven robust to noise and redundancy in the data, where supervised deep learning algorithms encounter problems Elad (2010). In the IQT context, low and high-quality dictionaries ($\mathbf{D}_\ell$, and $\mathbf{D}_h$ respectively) can be trained using a sparse representation model using pairs of low- and high-quality volumes. Subsequently, the sparse representation of a low-quality block, in terms of the low-quality dictionary $\mathbf{D}_\ell$, can be directly used to recover the corresponding high-quality block using the high-quality dictionary $\mathbf{D}_h$. As such, low-quality or high-quality volume patches are represented as a linear combinations of atoms drawn from a dictionary. SRep has been successfully applied to many other related inverse problems in image processing, such as denoising Li et al. (2012); Elad and Aharon (2006), restoration Zhang et al. (2014); Li et al. (2012), image quality assessment Liu et al. (2017, 2018, 2024, 2019), outlier or anomaly detection Eldaly (2018); Eldaly et al. (2019), image reconstruction Eldaly and Alexander (2024); Eldaly et al. (2025), and super resolution Yang et al. (2010). In a convex optimisation framework, training and testing samples are forced to follow the observation model of the imaging system on hand. Therefore, any new unseen test samples (either InD or OOD) will follow this model, which can avoid the "regression to the mean" problems observed with supervised regression models, often observed in OOD data.

On the other hand, in supervised deep learning, Dong et al. Dong et al. (2014) replaced the dictionary learning using sparse representation steps described above with a multilayered convolutional neural network to take advantage of the powerful capability of deep learning. As such, the low and high-quality dictionaries are implicitly acquired through network training. Various methods have been proposed to improve the performance of this approach such as in Kim et al. (2016); Lim et al. (2017); Tai et al. (2017); Zhang et al. (2018b). However, most of these studies, follow the same formality as in Dong et al. (2014) from a general perspective, where all the processes in the sparse-coding-based methods are replaced by a multilayered network. Recently, deep dictionary learning Tariyal et al. (2016) is proposed to take advantage of both transductive and inductive nature of dictionary learning and deep learning, respectively, and is very well suited where there is a scarcity of training data. While dictionary learning focuses on learning "basis" and "features" by matrix factorisation, deep learning focuses on extracting features via learning "weights" or "filter" in a greedy layer by layer fashion. Deep dictionary learning has been applied to various problems including recognition Tang et al. (2020); Sharma et al. (2017), image inpainting Deshpande et al. (2020), super resolution Huang and Dragotti (2018); Zhao et al. (2017), classification Majumdar and Singhal (2017); Majumdar and Ward (2017); Manjani et al. (2017), and load monitoring Singh and Majumdar (2017).

In this work, in contrast to existing IQT models in the literature, we propose two novel IQT algorithms, from which one is an example of unsupervised learning while the other is an example of blended supervised and unsupervised learning. The first approach is based on a sparse representation model and dictionary learning, which we call IQT-SRep. In this approach, low and high-quality dictionaries can be trained using a sparse representation model using pairs of low- and high-quality volumes. Subsequently, the sparse representation of a low-quality block, in terms of the low-quality dictionary, can be directly used to recover the corresponding high-quality block using the high-quality dictionary. The second approach is based on deep dictionary learning which we call IQT-DDL. This approach explicitly learns high-quality dictionary through network training. The main network predicts the high-quality dictionary coefficients, and the weighted sum of the dictionary atoms generates a high-quality output. This approach differs fundamentally from traditional deep-learning methods, which typically employ upsampling layers within the network. The upsampling process in our IQT approach is efficient since pre-generated high-quality dictionary serves as a magnifier during inference. Additionally, the main network no longer needs to retain pixel-level information in the high-quality space, enabling it to focus solely on predicting the dictionary coefficients. The main advantages of these two novel formulations are that they are robust to super resolve heavily OOD test data, and they are well suited where there is a scarcity of training data. We demonstrate the two models using experiments from a low-field MRI application and compare the results with the recently proposed state-of-the-art supervised deep learning approach Lin et al. (2022). As such, the main contributions of this paper can be summarised as follows.

1. We propose two new formulations of the IQT technique, from which one is an unsupervised learning based (IQT-SRep), and one is based on a combination of both supervised and unsupervised learning (IQT-DDL). Both of these formulations have never been previously applied to the IQT problem in literature.

2. The IQT-SRep approach is based on sparse representation and dictionary learning model and assumes that a given low- or high-quality volume patch can be represented as a linear combination of atoms drawn from a dictionary that is trained using training examples of pairs of low- and high-quality volume patches. This requires training of a pair of coupled dictionaries using a sparse representation model using pairs of low- and high-quality volumes.

3. The IQT-DDL approach is based on a combination of supervised and unsupervised learning using deep dictionary learning. This approach assumes that a given low- or high-quality volume patch can be represented as a non-linear combination of atoms drawn from a dictionary that is trained using training examples of pairs of low- and high-quality volume patches.

4. We demonstrate the performance of the model using experiments from a low-field MRI application, using both InD and OOD data, and compare with the state-of-the-art supervised deep learning IQT method, for low-field MRI enhancement.

The remaining sections of the paper are organised as follows. Section 2 formulates the problem of IQT using three learning techniques; the formulations that we propose here for IQT-SRep and IQT-DDL are described in detail, and finally, the supervised deep learning approach proposed in Lin et al. (2022) is briefly presented for comparison. Experiments conducted using a low-field MRI application synthesised using data from the human connectome project (HCP) are presented in Section 3. A general discussion is then presented in 4. Conclusions and future work are finally reported in Section 5.

# 2. Proposed Approaches

## 2.1 Image quality transfer using sparse representation and dictionary learning (IQT-SRep)

### 2.1.1 Imaging model

The IQT problem can be mathematically formulated as follows: Given an original vectorised high-quality volume $\mathbf{X} \in \mathbb{R}^M$, its corresponding low-quality version is denoted as $\mathbf{Y} \in \mathbb{R}^P$, where the relation between the two volumes can be modeled as

$$\mathbf{Y} = \mathbf{LHX} + \mathbf{W}, \tag{1}$$

where $\mathbf{H}$ is the matrix representing a linear blurring operator, $\mathbf{L}$ is the downsampling operator, and $\mathbf{W}$ stands for additive noise, modelling observation noise and model mismatch and is assumed to be a white Gaussian noise sequence. This equation states that $\mathbf{Y}$ is a blurred and down-sampled version of the original high-quality volume $\mathbf{X}$.

In IQT, the goal is to recover a high-quality volume $\hat{\mathbf{X}}$ given its blurred and down-sampled version $\mathbf{Y}$, such that $\hat{\mathbf{X}} \approx \mathbf{X}$. The problem of estimating $\mathbf{X}$ from $\mathbf{Y}$ in Eq. (1) is an ill-posed linear inverse problem (LIP), i.e., the matrix $\mathbf{LH}$ is singular and/or very ill-conditioned, since for a given low-quality input, infinitely many high-quality volumes satisfy the above equation. Consequently, this problem requires additional regularisation (or prior information from Bayesian perspective) in order to reduce uncertainties and improve estimation performance.

Figure 1 shows a schematic diagram to the IQT problem using a sparse representation model and dictionary learning. The proposed model consists of two separate stages. First, the coupled low-quality and high-quality dictionaries, $\mathbf{D}_\ell$ and $\mathbf{D}_h$ respectively, are constructed from training data set. Then, a reconstruction algorithm is applied to upscale a test low-quality volume to recover its high-quality version. This algorithm considers the patch-based sparse prior model to recover an estimate to the high-quality volume in a patch-by-patch basis. The following sections provide more details about the two stages mentioned above.

### 2.1.2 Joint dictionary construction

Constructing the high-quality and low-quality dictionaries requires a set of matched high- and low-quality volume patches. The training set is composed by a set of high-quality and the corresponding low-quality volumes. As proposed by Zeyde et al. (2010), the high-quality volumes are processed to obtain only the high-frequency information, whereas the intensity maps are used for the low-quality volumes. Each of the high- and low-quality volumes are then split into a set of 3D patches which are vectorised and training pairs are generated. Patches containing $> 80\%$ background voxels are excluded from the patch library. The coupled-dictionary training algorithm proposed by Zeyde et al. (2010) is then

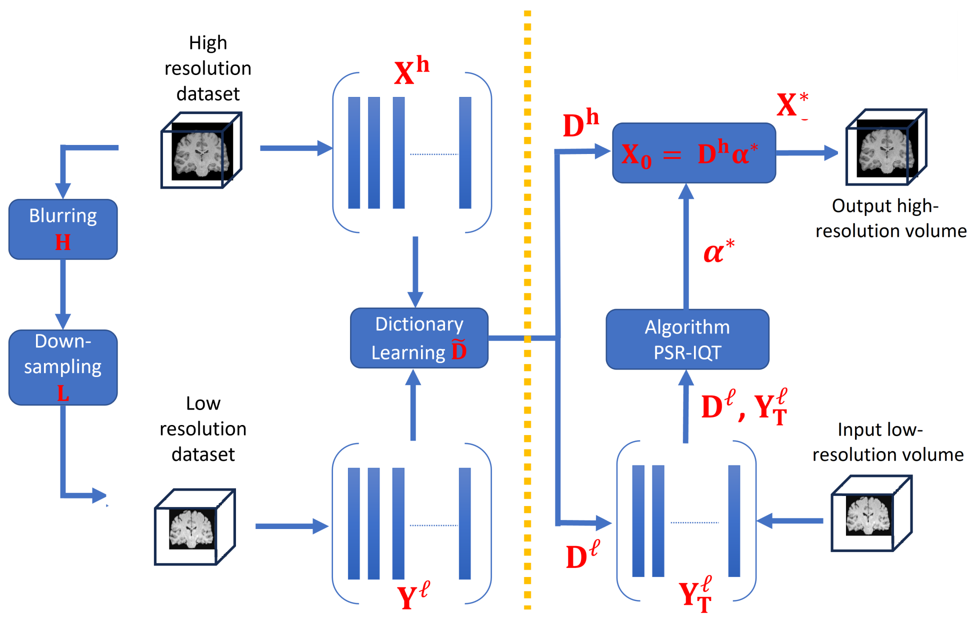


Figure 1: A schematic diagram of the proposed IQT approach using sparse representation and dictionary learning (IQT-SRep), where $\mathbf{D}_h$: High-quality dictionary, $\mathbf{D}_\ell$: Low-quality dictionary, $\mathbf{Y}$: Low-quality input volume, $\mathbf{X}_0$: Initial high-quality volume, $\lambda, \beta$ Regularisation parameters, $\sqrt[3]{m}$: Patch size, $p$: Number of pixel overlap, $s$: Scale, $\mathbf{y}$: A patch from the low-quality image $\mathbf{Y}$, $\mu$: Mean intensity of the patch $\mathbf{y}$, $\boldsymbol{\alpha}$: Sparse representation coefficients, $\boldsymbol{\alpha}^*$: Optimised sparse representation coefficients, $\mathbf{F}$: Transformation matrix, $\mathbf{x}$: High-quality patch, and $\mathbf{X}^*$: High-quality volume.

used in order to obtain the low- and high-quality dictionaries $\mathbf{D}_\ell$ and $\mathbf{D}_h$ respectively. For this local model, the two dictionaries $\mathbf{D}_h$ and $\mathbf{D}_\ell$ are trained such that they share the same sparse representations for each high- and low-quality volume patch pair. Finally, the dimensionality of $\mathbf{D}_\ell$ may be reduced to speed up the subsequent computations, given the intrinsic redundancy of the multi-scale edge analysis. For doing so, a Principal Component Analysis (PCA) is applied to this matrix, searching for a set of projection coefficients that represents at least 90% of the original variance. All patches are collected together to form the reduced low-quality dictionary $\mathbf{D}_\ell$, whereby the number of atoms in the dictionary has not changed.

#### 2.1.3 Patch-based sparsity prior model

The low-quality volume $\mathbf{Y}$ can be split into a set of overlapping 3D patches $\mathbf{y}$, each of size $\sqrt[3]{m} \times \sqrt[3]{m} \times \sqrt[3]{m}$. With the sparse generative model, each patch $\mathbf{y}$ can be represented by a linear combination of a few atoms drawn from a dictionary $\mathbf{D}_\ell$, which characterises the low-quality patches. This can be written as

$$\mathbf{y} = \mathbf{D}_\ell \boldsymbol{\alpha}_\ell, \tag{2}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^K$ is a sparse vector and $\|\boldsymbol{\alpha}\|_0 \ll K$. The corresponding high-quality patch $\mathbf{x}$, with size $\sqrt[3]{p} \times \sqrt[3]{p} \times \sqrt[3]{p}$, can be computed by again applying the following sparse generative model

$$\mathbf{x} = \mathbf{D}_h \boldsymbol{\alpha}_h. \tag{3}$$

From Eq. (2) and (3), it can be assumed that the sparse representation of a low-quality patch in terms of $\mathbf{D}_\ell$ can be directly used to recover the corresponding high-quality patch from $\mathbf{D}_h$, namely, that $\boldsymbol{\alpha}_\ell = \boldsymbol{\alpha}_h$. Therefore, the reconstructed high-quality image $\hat{\mathbf{X}}$ can be built by applying the sparse representation to each patch $\mathbf{y}$ in $\mathbf{Y}$ and then using the estimated $\boldsymbol{\alpha}$ with $\mathbf{D}_h$ to obtain each $\mathbf{x}$, which together form the image $\hat{\mathbf{X}}$.

#### 2.1.4 Local reconstruction by sparsity

The aim is to estimate a high-quality version $\tilde{\mathbf{X}}$ from a given low-quality volume $\mathbf{Y}$. Given a test low-quality volume, for each input low-quality patch $\mathbf{y}$, we find a sparse representation with respect to $\mathbf{D}_\ell$. The corresponding high-quality patch bases $\mathbf{D}_h$ will be combined according to these coefficients to generate the output high-quality patch $\mathbf{x}$. The problem of finding the sparsest representation of $\mathbf{y}$ can be formulated as

$$\underset{\boldsymbol{\alpha}}{\text{minimise}} \quad \frac{1}{2}\|\mathbf{F}\mathbf{D}_\ell\boldsymbol{\alpha} - \mathbf{F}\mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1, \tag{4}$$

where $\lambda$ balances sparsity of the solution and fidelity of the approximation to $\mathbf{y}$, and $\mathbf{F}$ is a linear feature extraction operator as in Zeyde et al. (2010). Given the optimal solution $\boldsymbol{\alpha}^*$ of Eq.(4), the high-quality patch $\mathbf{x}$ can be reconstructed as $\mathbf{x} = \mathbf{D}_h\boldsymbol{\alpha}^*$. This optimisation problem can be solved using the Basis Pursuit algorithm Chen and Donoho (1994).

The complete IQT process is summarised in Algorithm (1). In this algorithm, the input low-quality volume $\mathbf{Y}$ is up-sampled using bicubic interpolation to provide a preliminary high-resolution volume. For each cubic patch of size $\sqrt[3]{m} \times \sqrt[3]{m} \times \sqrt[3]{m}$ from the up-sampled volume, starting from the top left corner with an overlap $p$, the mean intensity $\mu$ of the patch is computed to ensure that the dictionary represents image textures rather than absolute intensities. The sparse representation $\boldsymbol{\alpha}^*$ of the patch is then obtained by solving an optimisation problem that minimises the difference between the transformed low-quality patch and its sparse representation in the low-quality dictionary $\mathbf{D}_\ell$, subject to a sparsity constraint controlled by the regularisation parameter $\lambda$. Using the high-quality dictionary $\mathbf{D}_h$ and the sparse coefficients $\boldsymbol{\alpha}^*$, a high-quality patch $\mathbf{x}$ is generated. This high-quality patch, with the mean intensity restored, is placed in the corresponding location in the initial high-quality volume $\mathbf{X}_0$. After processing all patches, the final high-quality volume $\mathbf{X}$ is obtained.

**Algorithm 1** IQT using patch-based sparse representation and dictionary learning (IQT-SRep)

1: **Input**: $\mathbf{D}_h, \mathbf{D}_\ell$ and $\mathbf{Y}$
2: **Initialise** $\mathbf{X}_0$, **Choose** Regularisation parameters $\lambda, \beta$, Patch-size $\sqrt[3]{m}$, pixel-overlap $p$ and scale $s$
3: Up sample the input low-quality volume using bicubic interpolation.
4: **For** each $\sqrt[3]{m} \times \sqrt[3]{m} \times \sqrt[3]{m}$ patch $\mathbf{y}$ from an image $\mathbf{Y}$, from top left corner of the volume, with an overlap $p$
- Compute: mean intensity $\mu$ of the patch $\mathbf{y}$
- Solve: $\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\text{minimise}} \ \ \frac{1}{2}\|\mathbf{F}\mathbf{D}_\ell\boldsymbol{\alpha} - \mathbf{F}\mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1$
- Generate the high-quality patch $\mathbf{x} = \mathbf{D}_h\boldsymbol{\alpha}^*$
- Place the high-quality patch $\mathbf{x} + \mu$ in the high-quality volume $\mathbf{X}_0$

5: **End**
6: **Output** High-quality volume $\mathbf{X} = \mathbf{X}_0$

### 2.2 Image quality transfer using deep dictionary learning (IQT-DDL)

The IQT using a deep dictionary learning model is composed of three main steps: constructing the high-quality dictionary $\mathbf{D}_H$, per-pixel prediction, and finally image reconstruction from patches. The high-quality dictionary $\mathbf{D}_H$ is generated from random noise input. The per-pixel predictor then estimates the coefficients of $\mathbf{D}_H$ for each pixel from a low-quality input. In the reconstruction phase, the high-quality image can be computed using the weighted sum of the elements (or atoms) of $\mathbf{D}_H$. In this work, we use L1 loss function to optimise the network $L = \frac{1}{M}\sum_{m=1}^{M}\|I_m^{gt} - \Theta(I_m^{lq})\|_1$, where $I_m^{lq}$ and $I_m^{gt}$ are low- and high-quality patches respectively, $M$ is the number of training pairs, and $\Theta(\cdot)$ represents a function of the IQT-DDL network. Figure 2 provides a schematic diagram of the proposed method. The following sections provide more details about each step.

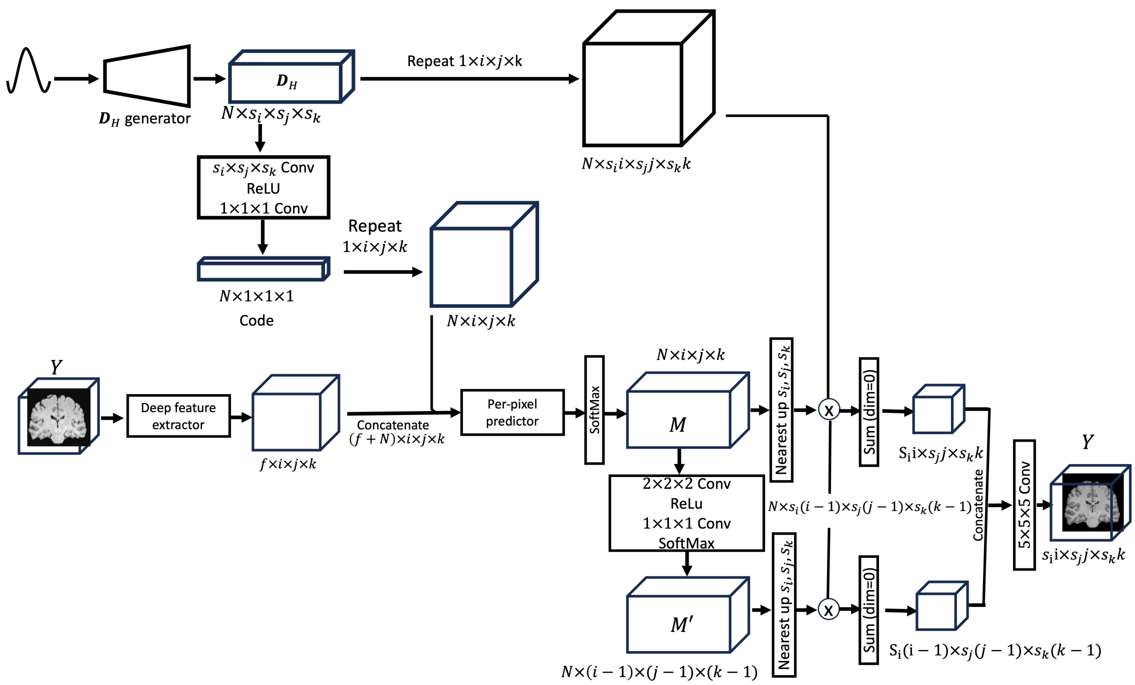


Figure 2: A schematic diagram of IQT using deep dictionary learning. Random noise generates the high-resolution dictionary $\mathbf{D}_H$. Then a per-pixel predictor takes as input a concatenation of an encoded code of $\mathbf{D}_H$ and an extracted feature. A final image based on $\mathbf{D}_H$ is then constructed using predictor output.

### 2.2.1 Construction of the high-quality dictionary $\mathbf{D}_H$

The high-quality dictionary $\mathbf{D}_H^{(N \times s_i \times s_j \times s_k)}$ is constructed from random noise using a standard Gaussian distribution, where $s_i$, $s_j$ and $s_k$ are up-scaling factors in $i$, $j$ and $k$ directions, and $N$ is the number of dictionary atoms. The high-quality dictionary $\mathbf{D}_H$ is then encoded by $s_i \times s_j \times s_k$ convolution with groups $N$, followed by ReLU Nair and Hinton (2010) and $1 \times 1 \times 1$ convolution. Each $N$ element of the resultant code $\mathbf{C}_H^{N \times 1 \times 1 \times 1}$ represents each $s_i \times s_j \times s_k$ atom as a scalar value. Note that low-quality dictionaries can be naturally replaced by convolutional operations, and therefore only $\mathbf{D}_H$ is constructed. The $\mathbf{D}_H$ generator has a tree-like structure, where the nodes consist of two $1 \times 1 \times 1$ convolutional layers with ReLU activation. The final layer has a Tanh activation followed by a pixel shuffling layer. To produce $N$ atoms, depth $d$ of the generator is determined as $d = \log_2(N)$.

### 2.2.2 Per-pixel Prediction

We use the UNet++ Zhou et al. (2018) as a deep feature extractor in Fig. 2, with depth of three, and a long skip connection is added. For an input image $I \in \mathbb{R}^{i \times j \times k}$, the deep feature extractor generates a tensor of size $f \times i \times j \times k$. The per-pixel predictor then takes as input a concatenation of the extracted feature and the expanded code of $\mathbf{D}_H$, such that $C_H^{N \times i \times j \times k} = R_{1 \times i \times j \times k}(C_H^{N \times 1 \times 1 \times 1})$, where $R_{a \times b \times c(\cdot)}$ denotes the $a \times b \times c$ repeat operations. The per-pixel predictor is composed of ten bottleneck residual blocks followed by a softmax function that computes the $N$ coefficients of $\mathbf{D}_H$ for each input pixel. Both the deep feature extractor and per-pixel predictor contain batch normalisation layers Loffe and Normalization (2014) before the ReLU activation. The resultant prediction map $M^{N \times i \times j \times k}$ is further convolved with a $2 \times 2 \times 2$ convolution layer to produce a complementary prediction map $M'^{N \times (i-1) \times (j-1) \times (k-1)}$, that compensates the patch boundaries when reconstructing the final output. The detail of the compensation mechanism is described in the next subsection.

### 2.2.3 Reconstruction

The prediction map $M^{N \times i \times j \times k}$ is upscaled to $N \times s_i i \times s_j j \times s_k k$ by nearest-neighbor interpolation, and the element-wise multiplication of that upscaled prediction map $U_{s_i s_j s_k}(M^{N \times i \times j \times k})$ with the expanded dictionary $R_{1 \times i \times j \times k}$ $(D_H^{N \times s_i \times s_j \times s_k})$ produces $N \times s_i i \times s_j j \times s_k k$ tensor $T$ consists of weighted atoms. The $U_{s_i, s_j, s_k}(\cdot)$ denotes $s_i \times s_j \times s_k$ nearest neighbor upsampling. Finally, the tensor $T$ is summed over the first dimension, producing the output $x$ as

$$x^{1 \times s_i i \times s_j j \times s_k k} = \sum_{n=0}^{N-1} T^{N \times s_i i \times s_j j \times s_k k}[n, :, :, :], \tag{5}$$

$$T^{N \times s_i i \times s_j j \times s_k k} = U_{s_i, s_j, s_k}(M^{N \times i \times j \times k}) \otimes R_{1 \times i \times j \times k}(D_H^{N \times s_i \times s_j \times s_k}). \tag{6}$$

The same sequence of operations is applied to the complementary prediction map to obtain the output $x'$. The final high-field prediction is obtained by centering $x$ and $x'$ on top of each other and concatenating the overlapping parts of the centered $x$ and $x'$, and applying a $5 \times 5 \times 5$ convolution. For non-overlapping parts, $x$ is simply used as the final output.

### 2.3 Image quality transfer using deep learning (IQT-DL)

A supervised learning IQT algorithm which was implemented using a deep learning framework (IQT-DL) is recently proposed Lin et al. (2022, 2019). This approach was used for IQT application in low-field MRI and showed superior performance compared to existing methods. The model is based on an anistropic U-Net trained on matched pairs of image patches from real high-field and synthetic low-field volumes generated by a stochastic decimation model which is presented in the Experiments section. This model considered the anisotropic U-Net architecture, which is an adaptation of the U-Net architecture to map input and output patches that differ in voxel dimension by the downsampling factor, $s$, in the slice direction. The main additions to the classic U-Net architecture are a bottleneck block, connecting corresponding levels of the contracting and expanding paths, and a residual core used to include more convolutional layers on each level. All convolution layers are activated by Rectified Linear Unit (ReLU) with Batch Normalisation (BN). The average voxel-wise mean square error over all patch pairs was used as a loss function. For more details and a block diagram of their proposed approach, see Lin et al. (2022).

## 3. Experiments

The performance of the proposed IQT-SRep and IQT-DDL approaches is demonstrated using a low-field MRI application, using both in-distribution (InD) and out-of distribution (OOD) datasets. The aim is to recover contrast enhanced and super-resolved images akin to those obtained using high field MRI scanners, standard in higher income countries, from low-field MR images form scanners still widely used in low-and-middle class income countries (LMICs). The proposed approaches are compared against the state-of-the-art supervised deep learning framework (IQT-DL) Lin et al. (2022, 2019), described in the previous section, to reveal both advantages and disadvantages of each of them. The main data set for training and testing is derived from the T1-weighted MRI images provided by the Human Connectome Projects (HCP), acquired on a 3 Tesla Siemens Connectome scanner Sotiropoulos et al. (2013a), with a 0.7-mm isotropic voxel. The repetition time (TR), echo time (TE), and inversion time (TI) for T1w are set to $2400, 2.14,$ and 1000 ms, respectively. We have chosen 65 subjects, from which 60 were used for training and 5 for testing. The training and testing datasets are synthesised using a stochastic low-field simulator described in Lin et al. (2022), the inputs of which are the signal-to-noise ratio (SNR) in gray matter (GM) and white matter (WM). The training data set is built using, for each synthetic volume, a randomly sampled SNR pair from the bivariate Gaussian distribution estimated from a real low-field MRI data set acquired in Nigeria Lin et al.

(2022). Three Low-field test datasets, five volumes each, are synthesised. Two test datasets are synthesised using parameters sampled from the same 2D Gaussian distribution used for the training set, and are called in-distribution data (InD1 and InD2). In particular, InD1 is synthesised with parameters using a Mahalanobis distance $< 1$, and InD2 with Mahalanobis distance $> 3$, with the constraint of having the SNR higher in WM than in GM, to keep the tissue contrast compatible with T1w. The simulation parameters of third data set are sampled from a distribution estimated from ultra-low field T1w images, and is called out-of-distribution (OOD) data set. Figure 3 shows a schematic diagram of both training and testing data structure, with the stochastic low-field image simulator for training and testing samples described below.

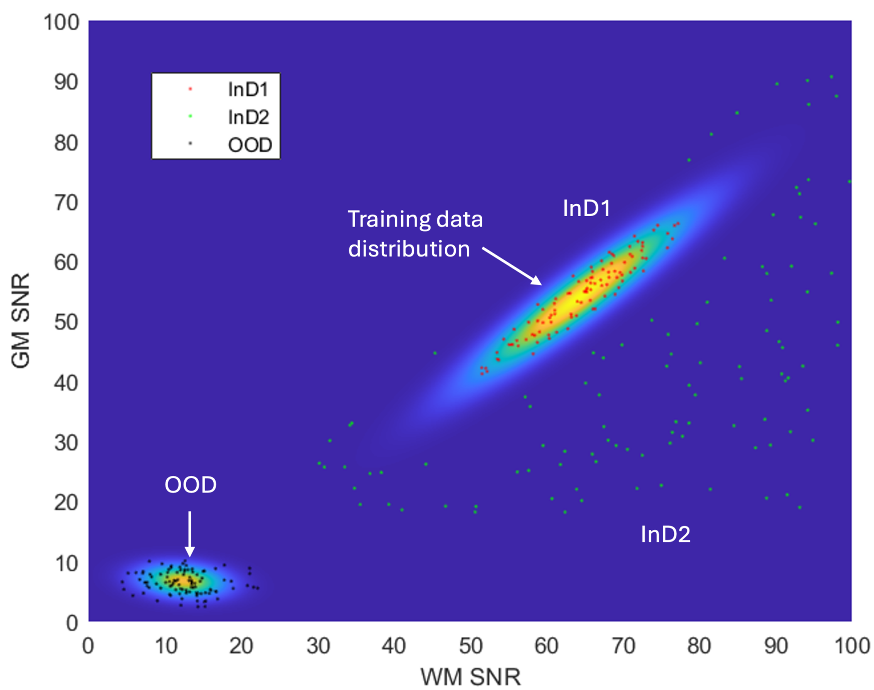


Figure 3: A schematic diagram of training and testing datasets. The two test in distribution datasets (InD1 and InD2) are synthesised using parameters sampled from the same 2D Gaussian distribution used for the training set. In particular, InD1 is synthesised with parameters using a Mahalanobis distance $< 1$, and InD2 with Mahalanobis distance $> 3$, with the constraint of having the SNR higher in WM than in GM, to keep the tissue contrast compatible with T1w. The out-of-distribution (OOD) data set is simulated using parameters sampled from a distribution estimated from ultra-low field T1w images.

### 3.1 Model training for IQT-SRep, IQT-DDL, and IQT-DL

Once the training set of matched low-and high-field pairs is composed as explained above, paired patches are obtained by cropping corresponding high-quality and synthetic low-quality volumes into patches at regularly spaced locations. Patches containing $> 80\%$ background voxels are excluded from the patch library. Training details of the IQT-SRep, IQT-DDL and IQT-DL models are presented below.

#### 3.1.1 IQT-SRep

The number of atoms and patch-sizes in dictionaries $\mathbf{D}_\ell$ and $\mathbf{D}_h$ has impact on two important aspects of the proposed IQT-SRep model; that are the reconstruction accuracy and reconstruction time. Larger dictionaries include more image patterns, and therefore more accurate super-resolved volumes. However, the drawbacks are the computational complexity of solving the optimisation problem and the longer time required for patch extraction. Following this, from an initial set of $100,000$ 3D-vectorised patches, we learned compact dictionaries of different atom numbers, including 150, 256, 512, 1024 and patch-sizes of $3 \times 3 \times 3$, $5 \times 5 \times 5$ and $7 \times 7 \times 7$. We first present those of 1024 atoms using $7 \times 7 \times 7$ patch-size, which provide best construction quality, and the effect of different atom number is presented afterwards.

#### 3.1.2 IQT-DDL

In this work, we adopt a model using different atoms numbers of 64 and 128 atoms. The number of filters of the models is adjusted according to the number of atoms. The scaling factors $s_i, s_j$, and $s_k$ are set to $s_i = 1, s_j = 1$, and $s_k = 4$. The network is trained using low-quality patch size of $32 \times 32 \times (32/s_k)$ with a mini-batch size of 32. Random flipping and rotation augmentation is applied to each training sample. An Adam optimiser Kingma and Ba (2014) with $\beta_1 = 0.9, \beta_2 = 0.999$, and $\epsilon = 10^{-8}$ is used. The learning rate of the network except for the $\mathbf{D}_H$ generator is initialised as $2e^{-4}$ and halved at $[200k, 300k, 350k, 375k]$. The total training iterations is $400k$. The learning rate of the $\mathbf{D}_H$ generator is initialised as $5e^{-3}$ and halved at $[50k, 100k, 200k, 300k, 350k]$. In addition, to stabilise training of the $\mathbf{D}_H$ generator, we randomly shuffle the order of output atoms for the first $1k$ iterations. The results of the 128 atoms dictionary are first presented, followed by a comparison with those of 64 atoms dictionary.

#### 3.1.3 IQT-DL

As in Lin et al. (2022), we use a default patch size of $32 \times 32 \times (32/s_k)$ and $32 \times 32 \times 32$, respectively for low-field and high-field volumes, and a step size of 8, 16, and $16/s_k$ along $x$-, $y$-, and $z$-directions, which provide best construction quality. Training model is then constructed using the training procedure explained in Section 2.3.

### 3.2 Testing

Each test volume is split into overlapping patches of size similar to that used for training in each model. The trained IQT-SRep, IQT-DDL and IQT-DL models described above are then applied to each of these patches to estimated the high-field volumes. The magnification factor $s_k$ for all models is set to 4. For the IQT-SRep model, in all experiments, the $\lambda$ parameter is set to 0.01. Slight variation of this parameter does not change the results significantly.

### 3.3 Evaluation

The quantitative measure used to assess the quality of the IQT algorithms presented in the previous section are the normalised root mean squared error (NRMSE), defined as

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{n=1}^{N}(\mathbf{x}_n - \hat{\mathbf{x}_n})^2}{N}}}{\text{Max}(\mathbf{x})}, \tag{7}$$

where $\mathbf{x}$ is the ground truth high-quality image, $\hat{\mathbf{x}}$ is the corresponding estimate from the low-field counterpart, and Max($\mathbf{x}$) is the maximum intensity of the ground truth high-field image $\mathbf{x}$, and structural similarity index measure (SSIM) which can be computed as in Wang et al. (2004).

### 3.4 Results

We utilise the proposed unsupervised learning IQT-SRep, the supervised deep learning IQT-DL and the blended learning IQT-DDL approaches to super resolve the testing datasets InD1, InD2 and OOD described above. Below, we show the quantitative and the qualitative performance, as well as the effect of changing different crucial parameters such as atom number in IQT-SRep and IQT-DDL approaches.

#### 3.4.1 Quantitative results

Table 1 provides NRMSE and SSIM results of InD1, InD2 and OOD using the three methods IQT-SRep, IQT-DDL and IQT-DL. We can observe that the supervised deep learning approach IQT-DL provides better results (lowest NRMSE and highest SSIM) using the in-distribution datasets (InD1 and InD2), compared to the unsupervised learning IQT-SRep algorithm, revealing that supervised learning is more robust for super-resolving images that follow the same distribution of the training data set compared to unsupervised learning. However, when testing using out-of distribution data that is different from the distribution of the training samples, the unsupervised learning approach IQT-SRep provides lower NRMSE and higher SSIM compared to the supervised deep learning model IQT-DL. This highlights the importance of unsupervised learning models since the full data distribution cannot be represented in the training data set. On the other hand, we can observe that the supervised deep learning model IQT-DL performs better (lower NRMSE and higher SSIM) than the blended supervised and unsupervised learning IQT-DDL approach using InD1, whereas the IQT-DDL provides better results using both InD2 and OOD datasets. This reveals the robustness of the blended learning IQT-DDL approach in super-resolving datasets differ from that the model was trained on, in addition to data that slightly deviates from InD1 but still part of the training samples.

### 3.4.2 Qualitative results

Figure 4 shows examples of coronal T1 weighted images from the HCP data set, corresponding to synthesised low-field images using InD1, InD2 and OOD, and results of IQT-SRep, IQT-DDL and IQT-DL. Figure 5 shows corresponding absolute error maps between high-quality ground truth images and corresponding low-quality images, and results of IQT-SRep, IQT-DDL and IQT-DL. Moreover, the binary maps of regions (in red label) where the IQT-SRep and IQT-DDL models provide closer estimates to ground truth high-quality images compared to IQT-DL are also presented. The qualitative results in general follow the same behaviour of the quantitative results described earlier: although IQT-DL provides better visual results of brain structure compared to IQT-SRep using InD1, and InD2, the IQT-SRep model shows better visual results using OOD data compared to IQT-DL. This is clearer in the absolute error maps in Fig. 5, between high-quality ground truth images and results of both IQT-DL and IQT-SRep, and in the binary maps where there are more image regions where IQT-SRep and IQT-DDL performs better than IQT-DL. This implies that the IQT-SRep approach is more robust for image enhancement using out-of-distribution data, which are created using a different distribution to that of the training samples mimicking real-world examples. Moreover, the IQT-SRep approach provides smoother outputs compared to that of the IQT-DL approach where artifacts arising from patch construction are very obvious. On the other hand, the blended learning IQT-DDL approach provides better visual results than the supervised deep learning approach IQT-DL using InD2 and OOD datasets. This is also clear in the absolute error maps and in the binary maps where there are more image regions where IQT-DDL performs better than IQT-DL in Fig. 5. On the other hand, while in this work we process data volumes by splitting them into overlapping patches, the proposed approaches ensure that information from the borders of each patch is preserved and integrated into the subsequent patches, thereby there is no information loss and the continuity of image features across the entire volume is maintained. Moreover, we synthesise the low-quality volumes from the high-quality ones, which ensures that there are no pixel alignment problems, as both low-quality and high-quality volume pairs are inherently aligned during the synthesis.

To summarise, the blended learning IQT-DDL approach provides best visual results compared to the supervised deep learning IQT-DL and the unsupervised leaning IQT-SRep approaches using both InD2 and OOD datasets, whereas the unsupervised learning approach IQT-SRep provides better visual results than the supervised deep learning IQT-DL approach using OOD which is generated using a different distribution to that of the training samples. There are widespread regions where the errors are lower for IQT-SRep

Table 1: Normalised root mean squared error (NRMSE), and structural similarity index measure (SSIM) using in-distribution data (InD1), and (InD2), and out-of-distribution data (OOD). Best results are highlighted in bold font, and second best are underlined.

| | **Interpolation** | | **IQT-SRep** | | **IQT-DDL** | | **IQT-DL** | |
|---|---|---|---|---|---|---|---|---|
| | NRMSE | SSIM | NRMSE | SSIM | NRMSE | SSIM | NRMSE | SSIM |
| **InD1** | 0.257 | 0.698 | 0.240 | 0.711 | 0.126 | 0.792 | **0.096** | **0.869** |
| **InD2** | 0.328 | 0.612 | 0.319 | 0.641 | **0.238** | **0.732** | 0.258 | 0.724 |
| **OOD** | 0.469 | 0.585 | 0.450 | 0.632 | **0.435** | **0.642** | 0.455 | 0.630 |

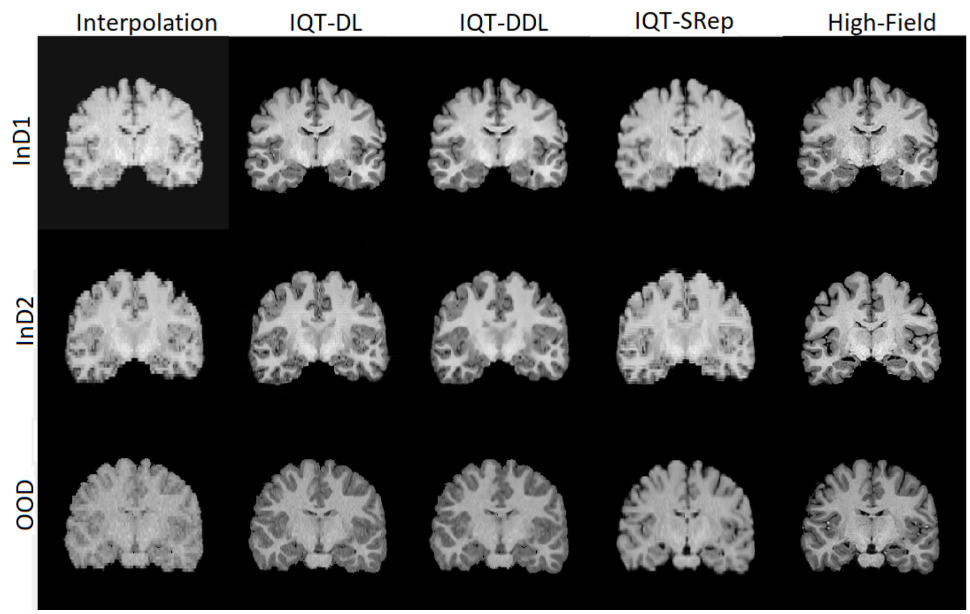


Figure 4: Results using the HCP data set on coronal direction of the three different data distributions InD1, InD2 and OOD (rows) using IQT-SRep, IQT-DDL and IQT-DL. First column shows interpolated low-field image, second to forth columns show image estimate using IQT-DL, IQT-DDL and IQT-SRep, respectively, and fifth column shows original high-field image.

and IQT-DDL compared to IQT-DL highlighting bias in the regression model estimates, which both IQT-SRep and IQT-DDL can avoid.

#### 3.4.3 Effect of atom number and output patch size

Now, we evaluate the effect of atom number and patch size on both approaches. From the sampled 100,000 image patch pairs, and for the IQT-SRep approach, we train four dictionaries of size 150, 256, 512, 1024, and use each to estimate the high-field image

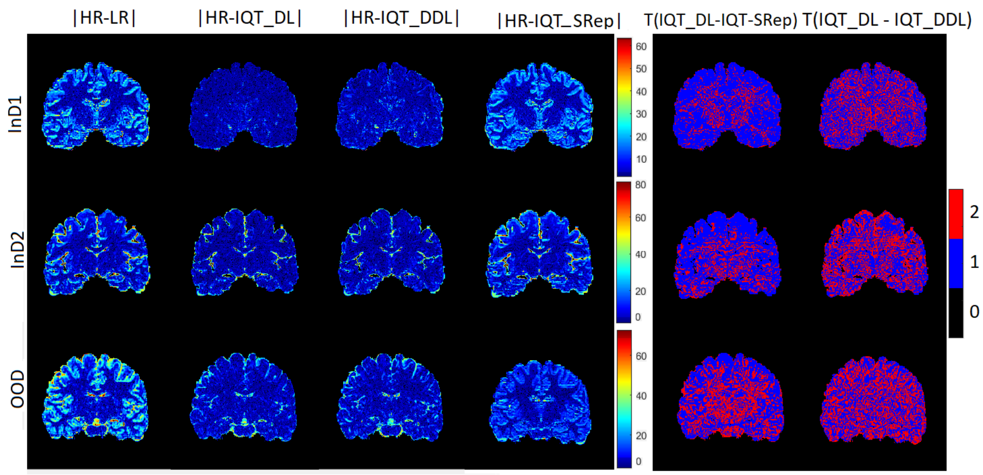


Figure 5: Absolute errors for results in Figure 4, between gold-standard high-field image (Column 5 of Figure 4), and Column 1: corresponding low-quality image, Column 2: IQT-DL, Column 3: IQT-DDL, and Column 4: IQT-SRep. Columns 5 and 6 show binary maps of regions (in red label) where the IQT-SRep and IQT-DDL, respectively provide closer estimates to the gold-standard high-field images compared to IQT-DL.

from the low-field counterpart. Moreover, for the IQT-DDL approach, in order to assess the performance of the algorithm using different atoms numbers, we construct dictionaries using atom number of 64, in addition to that of 128 whose results are presented in the previous section. Table 2 shows NRMSE of the IQT-SRep and IQT-DDL approaches using different atom number using the three testing datasets InD1, InD2 and OOD. We can observe that in general, as atom number increases, construction quality improves (NRMSE decreases and SSIM increases), but saturates for atom number higher than 512 for the IQT-SRep approach. For the IQT-SRep approach, all tested atom numbers still provide better construction results using OOD data as compared to the IQT-DL approach. On the other hand, we tested several patch sizes for both the IQT-SRep and IQT-DDL models. Specifically, for the IQT-SRep model, we tested patch sizes of P3: $3 \times 3 \times 3$, P5: $5 \times 5 \times 5$, and P7: $7 \times 7 \times 7$ using a dictionary size of 1024 (which provides the best results). For the IQT-DDL model, we tested patch sizes of P16: $16 \times 16 \times 16$, P32: $32 \times 32 \times 32$, and P48: $48 \times 48 \times 48$ using a dictionary size of 128 (which provides the best results). Table 3 provides the NRMSE and SSIM for two in-distribution datasets (InD1 and InD2) and one out-of-distribution (OOD) dataset. We can observe that for the IQT-SRep model, the reconstruction results improves (NRMSE decreases and SSIM increases) as the patch size increases. Conversely, for the IQT-DDL model, a patch size of P32: $32 \times 32 \times 32$ outperforms both P16: $16 \times 16 \times 16$ and P48: $48 \times 48 \times 48$, indicating that it is a good operating point, balancing structural information content with the ability to learn and generalise from a finite training set. Fig. 6 shows an example of super resolved images using the OOD data set using dictionaries of different sizes at patch sizes providing best results (P7 for IQT-SRep and P32 for IQT-DDL). While there are no substantial visual differences, we can

Table 2: Effect of atom number for the IQT-SRep and IQT-DDL methods: NRMSE and SSIM using in-distribution data (InD1), and (InD2), and out-of-distribution data (OOD).

| | | **IQT-SRep** | | | | **IQT-DDL** | |
|---|---|---|---|---|---|---|---|
| | | **D150** | **D256** | **D512** | **D1024** | **D64** | **D128** |
| **InD1** | NRMSE | 0.243 | 0.242 | 0.240 | 0.240 | 0.128 | 0.126 |
| | SSIM | 0.704 | 0.705 | 0.706 | 0.706 | 0.791 | 0.792 |
| **InD2** | NRMSE | 0.322 | 0.321 | 0.319 | 0.319 | 0.240 | 0.238 |
| | SSIM | 0.639 | 0.640 | 0.641 | 0.641 | 0.731 | 0.732 |
| **OOD** | NRMSE | 0.452 | 0.451 | 0.450 | 0.450 | 0.437 | 0.435 |
| | SSIM | 0.630 | 0.631 | 0.632 | 0.632 | 0.641 | 0.642 |

observe in the binary error maps that the number red pixels (improvement over interpolated low-field image) gradually increase with larger dictionaries until saturation for dictionary size of 1024. In terms of computation time, the IQT-SRep algorithm is implemented in MATLAB and the experiments are carried out on a laptop with a 2.8 GHz processor CPU, with 16 GB of RAM, under Microsoft Windows 10. Dictionary construction times ranges from $\sim$ 25 min for a 150-size dictionary to $\sim$ 80 min for a 1024-size dictionary. During the testing, in terms of test image reconstruction time, the computation is approximately linear to the size of the dictionary, that larger dictionaries will result in heavier computation. For example, smaller dictionaries, such as those with 150 atoms, yield reconstructions in an average time of $\sim$ 7 min, while larger dictionaries, such as those with 1024 atoms, yielded image reconstructions in an average time of 50 min. On the other hand, for the IQT-DDL algorithm, as shown in Table 2, the NRMSE is slightly lower using dictionary with atom number of 128 compared to that of 64 atoms for all testing datasets InD1, InD2 and OOD, as it retains more image patterns. Fig. 6 shows visual results of the IQT-DDL approach using an OOD example. Similar to the IQT-SRep approach, while there is no substantial visual difference, we indeed observe the increase in more super-resolved pixels (in red) in the binary error map images compared to the interpolated low-field image for atom number of 128 compared to that of 64. In terms of computation time, the IQT-DDL algorithm is implemented in PyTorch, and the testing construction time ranges from $\sim$ 4 to 7 min for atom numbers of 64 to 128, respectively.

## 4. Discussion

This work introduced two novel IQT approaches. To the best of our knowledge, it is the first time in the literature that an unsupervised learning and a blended supervised and unsupervised learning frameworks are considered for IQT. These approaches are introduced to highlight biased estimates that can result from supervised learning approaches, such as supervised deep learning, especially using out-of-distribution data. The main advantages of these two novel formulations are that they tend to avoid biased estimates using out-of-distribution test data, and are very well suited where there is a scarcity of training

Table 3: Effect of output patch size for the IQT-SRep (using D1024) and IQT-DDL (using D128) methods: NRMSE and SSIM using in-distribution data (InD1), and (InD2), and out-of-distribution data (OOD).

| | | IQT-SRep | | | IQT-DDL | | |
|---|---|---|---|---|---|---|---|
| | | **P3** | **P5** | **P7** | **P16** | **P32** | **P48** |
| **InD1** | NRMSE | 0.250 | 0.244 | 0.240 | 0.133 | 0.126 | 0.129 |
| | SSIM | 0.671 | 0.702 | 0.706 | 0.768 | 0.792 | 0.785 |
| **InD2** | NRMSE | 0.325 | 0.322 | 0.319 | 0.244 | 0.238 | 0.237 |
| | SSIM | 0.635 | 0.639 | 0.641 | 0.725 | 0.732 | 0.729 |
| **OOD** | NRMSE | 0.459 | 0.455 | 0.450 | 0.440 | 0.435 | 0.437 |
| | SSIM | 0.625 | 0.628 | 0.632 | 0.635 | 0.642 | 0.638 |

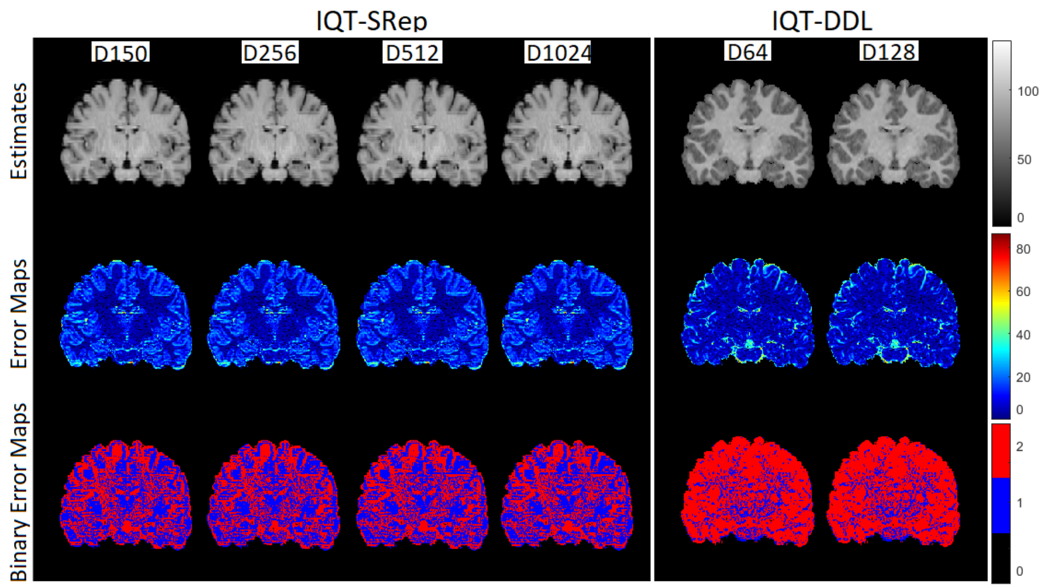


Figure 6: Effect of atom number for the IQT-SRep and IQT-DDL approaches using the OOD example in Fig. 4. Row 1: Image estimates using different atom numbers (from left to right: 150, 256, 512 and 1024 atoms (for IQT-SRep), and 64 and 128 atoms (for IQT-DDL). Row 2: Absolute difference error maps between the high quality image and each estimate in Row 1, and Row 3: Binary maps between interpolated low-field images and estimates in Row 1, where we can observe, as atom number increases, slightly more image regions (in red) are observed, until saturation from atom numbers of 512 to 1024 for IQT-SRep.

data. The first approach is based on a sparse representation and dictionary learning model, which trains two dictionaries using a sparse representation model from pairs of low- and high-quality volumes, whereas the second is based on a deep dictionary learning approach which explicitly learns high-resolution dictionary to upscale the input volume as in the sparse-coding-based methods, while the entire network, including high dictionary generator, is simultaneously optimised to take full advantage of deep learning methods. The performance of both approaches is demonstrated using a low-field MRI application, and compared against state-of-the-art supervised deep learning algorithm using both in-distribution and out-of-distribution datasets. Although supervised deep learning approach showed a superior performance using an in-distribution data set, one disadvantage of such class of methods is that their performance is degraded for images with a different contrast than in the training data set (OOD data). The results presented in the previous section show that the sparsity prior for image patches in the IQT-SRep and approach is effective in regularising the ill-posed IQT problem leading to good performance using out-of-distribution data compared to the IQT-DL approach. In these results, the dictionary size is fixed be 1024. Obviously, larger dictionaries retain more expressive patterns to the volumes of the trained data set, thus, yield more accurate approximation to the sparsity optimisation problem during the testing phase. However, this comes at the expense of increasing the computation cost. On the other hand, the blended learning IQT-DDL approach shows that the upsampling process is efficient because the main network does not need to maintain the information of the processed image at the pixel level in high-quality image space. Therefore, the network can concentrate only on predicting the coefficients of the high-quality dictionary yielding better performance using the InD2 and OOD datasets compared to IQT-DL and IQT-SRep. Extensive experiments show that sparse representation using dictionary learning and the deep dictionary learning approaches are more robust in super-resolving out-of-distribution test images compared to supervised deep learning. On the other hand, unsupervised learning is more robust to noise and redundancy in the data compared to supervised learning. Precisely, in a convex optimisation framework, training and testing samples are forced to follow the observation model of the imaging system on hand, and therefore, any new unseen test samples will follow this model, which can avoid the "regression to the mean" problems observed with supervised regression models. The biased estimates produced using the IQT-DL model likely arise because these image regions are under-represented in the training data, and thus the model is under-fit, which further adds bias in estimates. Although other unsupervised approaches, in particular deep unsupervised learning, might produce better results, the proposed approach highlight the problem and provide a baseline potential solution. It is worth pointing out that the proposed methods performed slightly better than the supervised approach only for data that were quite significantly different from the training dataset; this might be a not very relevant scenario for simulation-based training approaches, as most of the current IQT implementations are, as it would be much more advantageous to adapt the simulation parameters for the specific application and train an ad hoc supervised model than to adopt an unsupervised approach. However, there are several situations in which test images may have features that are difficult to simulate or predict, e.g. pathological alterations or artifacts. Some applications may also require very complex models that would be impractical to retrain for every single applications. In both these cases, it is critical to have a model that is robust enough to OOD data. Furthermore, as already mentioned, this was a proof-of-concept study considering relatively simple supervised approaches. More advanced methods will be investigated in the future and are expected to provide a more significant advantage compared to supervised baselines.

## 5. Conclusion and Future Work

In this work, we introduced two novel formulations of the IQT problem, which use an unsupervised learning framework, and a blended supervised and unsupervised learning, respectively. The unsupervised learning approach considers a sparse representation and dictionary learning model, whereas the combination of supervised and unsupervised learning approach is based on deep dictionary learning. The two models are evaluated using a low-field magnetic resonance imaging application aiming to recover high-quality images akin to those obtained from high-field scanners. Experiments comparing the proposed approaches against state-of-the-art deep learning IQT method identified that the two novel formulations of the IQT problem can avoid bias associated with supervised methods when tested using out-of-distribution data that differs from the distribution of the data the model was trained on. This highlights the potential benefit of these novel paradigms for IQT. Future work involves demonstrating the performance of the approach using real low-field MRI data and providing uncertainty bounds to the estimates, as well as extension to deep unsupervised methods that can combine the high fidelity appearance of supervised deep learning approaches to image enhancement with the reduced bias provided by unsupervised learning. The reduction of bias is an important step in the deployment of learning based methods for image enhancement, itself a vital component in the realisation of the potential of emerging low-field and portable MRI systems particularly for deployment in regions where accessibility is currently low.

## Acknowledgments

The 3T T1-weighted and T2-weighted images were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research, United States; and by the McDonnell Center for Systems Neuroscience at Washington University, United States. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

## Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The IQT models are trained and evaluated on the publicly available high-field MRI datasets, i.e. the HCP data set for T1w and T2w images Sotiropoulos et al. (2013b).

## References


Daniel C Alexander, Darko Zikic, Jiaying Zhang, Hui Zhang, and Antonio Criminisi. Image quality transfer via random forest regression: applications in diffusion mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 225–232. Springer, 2014.

Daniel C Alexander, Darko Zikic, Aurobrata Ghosh, Ryutaro Tanno, Viktor Wottschel, Jiaying Zhang, Enrico Kaden, Tim B Dyrby, Stamatios N Sotiropoulos, Hui Zhang, et al. Image quality transfer and applications in diffusion mri. *NeuroImage*, 152:283–298, 2017.

Udunna C Anazodo, Jinggang J Ng, Boaz Ehiogu, Johnes Obungoloch, Abiodun Fatade, Henk JMM Mutsaerts, Mario Forjaz Secca, Mamadou Diop, Abayomi Opadele, Dianel C Alexander, et al. A framework for advancing sustainable mri access in africa. *medRxiv*, 2022.

Stefano B Blumberg, Ryutaro Tanno, Iasonas Kokkinos, and Daniel C Alexander. Deeper image quality transfer: Training low-memory neural networks for 3d images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 118–125. Springer, 2018.

Shaobing Chen and David Donoho. Basis pursuit. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 41–44. IEEE, 1994.

S Deshpande, M Girish Chandra, and P Balamurali. Deep dictionary learning for inpainting. In *Computer Vision, Pattern Recognition, Image Processing, and Graphics: 7th National Conference, NCVPRIPG 2019, Hubballi, India, December 22–24, 2019, Revised Selected Papers*, volume 1249, page 79. Springer Nature, 2020.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014.

Michael Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*, volume 2. Springer, 2010.

Michael Elad and Michal Aharon. Image denoising via learned dictionaries and sparse representation. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 895–900. IEEE, 2006.

Ahmed Karam Eldaly. *Bayesian Image Restoration and Bacteria Detection in Optical Endomicroscopy*. PhD thesis, Heriot-Watt University, Edinburgh, United Kingdom, 2018.

Ahmed Karam Eldaly and Daniel C. Alexander. Bayesian magnetic resonance image reconstruction and uncertainty quantification. In *ISMRM & SMRT Virtual Conference & Exhibition*, Singapore, May 2024.

Ahmed Karam Eldaly, Yoann Altmann, Ahsan Akram, Antonios Perperidis, Kevin Dhaliwal, and Steve McLaughlin. Patch-based sparse representation for bacterial detection. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 657–661. IEEE, 2019.

Ahmed Karam Eldaly, Matteo Figini, and Daniel Alexander. Bayesian magnetic resonance joint image reconstruction and uncertainty quantification using sparsity prior models. *Submitted to IEEE transactions on Computational imaging*, 2025.

Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1604–1613, 2019.

Noemi G Gyori, Marco Palombo, Christopher A Clark, Hui Zhang, and Daniel C Alexander. Training data distribution significantly impacts the estimation of tissue microstructure with machine learning. *Magnetic resonance in medicine*, 87(2):932–947, 2022.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

Jun-Jie Huang and Pier Luigi Dragotti. A deep dictionary model for image super-resolution. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6777–6781. IEEE, 2018.

Juan E Iglesias, Benjamin Billot, Yaël Balbastre, Colin Magdamo, Steven E Arnold, Sudeshna Das, Brian L Edlow, Daniel C Alexander, Polina Golland, and Bruce Fischl. Synthsr: A public ai tool to turn heterogeneous clinical brain scans into high-resolution t1-weighted images for 3d morphometry. *Science advances*, 9(5):eadd3607, 2023.

Juan Eugenio Iglesias, Benjamin Billot, Yaël Balbastre, Azadeh Tabari, John Conklin, R Gilberto González, Daniel C Alexander, Polina Golland, Brian L Edlow, Bruce Fischl, et al. Joint super-resolution and synthesis of 1 mm isotropic mp-rage volumes from clinical mri exams with scans of different orientation, resolution and contrast. *NeuroImage*, 237: 118206, 2021.

Juan Eugenio Iglesias, Riana Schleicher, Sonia Laguna, Benjamin Billot, Pamela Schaefer, Brenna McKaig, Joshua N Goldstein, Kevin N Sheth, Matthew S Rosen, and W Taylor

Kimberly. Quantitative brain morphometry of portable low-field-strength mri using super-resolution machine learning. *Radiology*, 306(3):e220522, 2022.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.

Seunghoi Kim, Henry FJ Tregidgo, Ahmed K Eldaly, Matteo Figini, and Daniel C Alexander. A 3d conditional diffusion model for image quality transfer–an application to low-field mri. *arXiv preprint arXiv:2311.06631*, 2023.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

Vick Lau, Linfang Xiao, Yujiao Zhao, Shi Su, Ye Ding, Christopher Man, Xunda Wang, Anderson Tsang, Peng Cao, Gary KK Lau, et al. Pushing the limits of low-cost ultralow-field mri by dual-acquisition deep learning 3d superresolution. *Magnetic Resonance in Medicine*, 2023.

Shutao Li, Haitao Yin, and Leyuan Fang. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Transactions on biomedical engineering*, 59(12):3450–3459, 2012.

Yixiao Li, Xiaoyuan Yang, Jun Fu, Guanghui Yue, and Wei Zhou. Deep bi-directional attention network for image super-resolution quality assessment. *arXiv preprint arXiv:2403.10406*, 2024.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

Hongxiang Lin, Matteo Figini, Ryutaro Tanno, Stefano B Blumberg, Enrico Kaden, Godwin Ogbole, Biobele J Brown, Felice D'Arco, David W Carmichael, Ikeoluwa Lagunju, et al. Deep learning for low-field to high-field mr: image quality transfer with probabilistic decimation simulator. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 58–70. Springer, 2019.

Hongxiang Lin, Yukun Zhou, Paddy J Slator, and Daniel C Alexander. Generalised super resolution for quantitative mri using self-supervised mixture of experts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 44–54. Springer, 2021.

Hongxiang Lin, Matteo Figini, Felice D'Arco, Godwin Ogbole, Ryutaro Tanno, Stefano Blumberg, Lisa Ronan, Biobele J. Brown, David W. Carmichael, Ikeoluwa Lagunju, Judith Helen Cross, Delmiro Fernandez-Reyes, and Daniel C. Alexander. Low-field magnetic resonance image enhancement via stochastic image quality transfer. *Medical Image Analysis*, 2022.

Yutao Liu, Guangtao Zhai, Ke Gu, Xianming Liu, Debin Zhao, and Wen Gao. Reduced-reference image quality assessment in free-energy principle and sparse representation. *IEEE Transactions on Multimedia*, 20(2):379–391, 2017.

Yutao Liu, Ke Gu, Shiqi Wang, Debin Zhao, and Wen Gao. Blind quality assessment of camera images based on low-level and high-level statistical features. *IEEE Transactions on Multimedia*, 21(1):135–146, 2018.

Yutao Liu, Ke Gu, Yongbing Zhang, Xiu Li, Guangtao Zhai, Debin Zhao, and Wen Gao. Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(4):929–943, 2019.

Yutao Liu, Baochao Zhang, Runze Hu, Ke Gu, Guangtao Zhai, and Junyu Dong. Underwater image quality assessment: Benchmark database and objective method. *IEEE Transactions on Multimedia*, 2024.

S Loffe and CSB Normalization. Accelerating deep network training by reducing internal covariate shift. *arXiv*, 2014.

Angshul Majumdar and Vanika Singhal. Noisy deep dictionary learning: Application to alzheimer's disease classification. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2679–2683. IEEE, 2017.

Angshul Majumdar and Rabab Ward. Robust greedy deep dictionary learning for ecg arrhythmia classification. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 4400–4407. IEEE, 2017.

Ishan Manjani, Snigdha Tariyal, Mayank Vatsa, Richa Singh, and Angshul Majumdar. Detecting silicone mask-based presentation attack via deep dictionary learning. *IEEE Transactions on Information Forensics and Security*, 12(7):1713–1723, 2017.

Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464): 447–453, 2019.

Pulkit Sharma, Vinayak Abrol, and Anil Kumar Sao. Deep-sparse-representation-based features for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(11):2162–2175, 2017.

Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3118–3126, 2018.

Shikha Singh and Angshul Majumdar. Deep sparse coding for non–intrusive load monitoring. *IEEE Transactions on Smart Grid*, 9(5):4669–4678, 2017.

Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3516–3525, 2020.

Stamatios N. Sotiropoulos, Saad Jbabdi, Junqian Xu, Jesper L. Andersson, Steen Moeller, Edward J. Auerbach, Matthew F. Glasser, Moises Hernandez, Guillermo Sapiro, Mark Jenkinson, David A. Feinberg, Essa Yacoub, Christophe Lenglet, David C. Van Essen, Kamil Ugurbil, and Timothy E.J. Behrens. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage*, 80:125–143, oct 2013a. ISSN 10538119.

Stamatios N Sotiropoulos, Saad Jbabdi, Junqian Xu, Jesper L Andersson, Steen Moeller, Edward J Auerbach, Matthew F Glasser, Moises Hernandez, Guillermo Sapiro, Mark Jenkinson, et al. Advances in diffusion mri acquisition and processing in the human connectome project. *Neuroimage*, 80:125–143, 2013b.

Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017.

Hao Tang, Hong Liu, Wei Xiao, and Nicu Sebe. When dictionary learning meets deep learning: Deep dictionary learning and coding network for image recognition with limited data. *IEEE transactions on neural networks and learning systems*, 32(5):2129–2141, 2020.

Ryutaro Tanno, Daniel E Worrall, Enrico Kaden, Aurobrata Ghosh, Francesco Grussu, Alberto Bizzi, Stamatios N Sotiropoulos, Antonio Criminisi, and Daniel C Alexander. Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion mri. *NeuroImage*, 225:117366, 2021.

Snigdha Tariyal, Angshul Majumdar, Richa Singh, and Mayank Vatsa. Deep dictionary learning. *IEEE Access*, 4:10096–10109, 2016.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Yu-Syuan Xu, Shou-Yao Roy Tseng, Yu Tseng, Hsien-Kai Kuo, and Yi-Min Tsai. Unified dynamic convolutional network for super-resolution with variational degradations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12496–12505, 2020.

Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.

Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.

Jian Zhang, Debin Zhao, and Wen Gao. Group-based sparse representation for image restoration. *IEEE transactions on image processing*, 23(8):3336–3351, 2014.

Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3262–3271, 2018a.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018b.

Liling Zhao, Quansen Sun, and Zelin Zhang. Single image super-resolution based on deep learning features and dictionary model. *IEEE Access*, 5:17126–17135, 2017.

Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2433–2443, 2019.

Wei Zhou, Qiuping Jiang, Yuwang Wang, Zhibo Chen, and Weiping Li. Blind quality assessment for image superresolution using deep two-stream convolutional networks. *Information Sciences*, 528:205–218, 2020.

Wei Zhou, Zhou Wang, and Zhibo Chen. Image super-resolution quality assessment: Structural fidelity versus statistical naturalness. In *2021 13th International conference on quality of multimedia experience (QoMEX)*, pages 61–64. IEEE, 2021.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.