Appendix

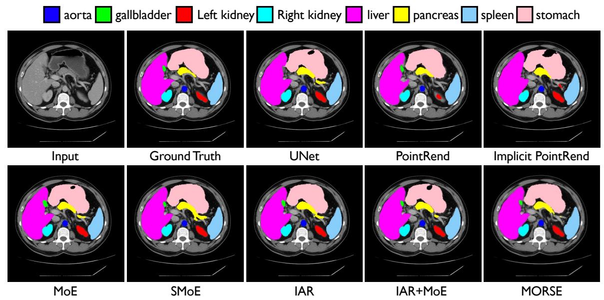

Fig. 2. Visualization results on Synpase. MORSE yields more accurate predictions, especially for small regions.

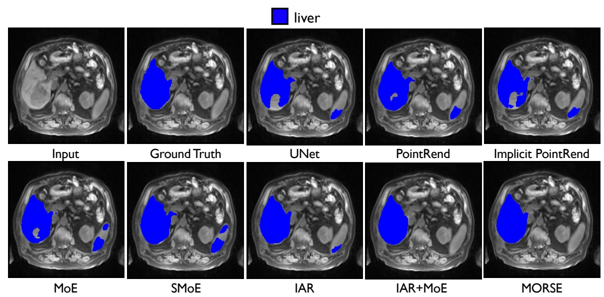

Fig. 3. Visualization results on MP-MRI. MORSE outputs more accurate segmentation results, especially for boundaries.

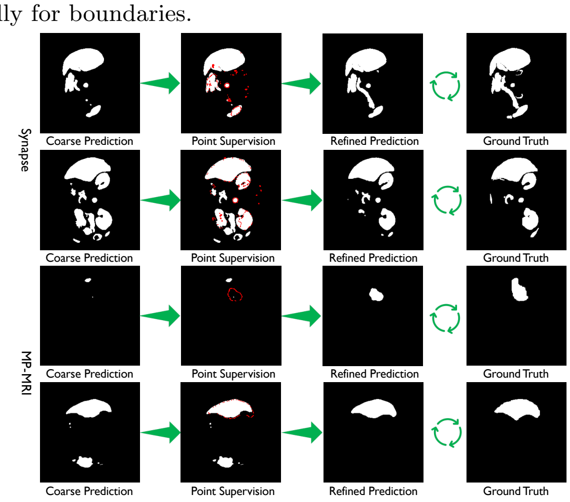

Fig. 4. Visualization of sampled point location. As is shown, IAR significantly improves the segmentation quality.

A Theoretical Analysis

In this section, we theoretically analyze the expressiveness of MORSE, and demonstrate the approximation power of INR features combined with MLPs.

We first introduce the kernel method which is commonly used in analyzing neural networks. We define the kernel function κ: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, which is a symmetric function measuring the similarity between two vectors in $\mathbb{R}^d$. Kernel functions are used to approximate unknown functions from data. Specifically, given a training dataset $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$ where $\mathbf{y}_i = \mu(\mathbf{x}_i)$ for some unknown function $\mu(\cdot)$, an estimate $\hat{\mu}$ can be constructed using kernel function as $\hat{\mu}(\mathbf{x}) = \sum_{i=1}^n (\mathbf{K}^{-1}\mathbf{Y})_i \kappa(\mathbf{x}_i, \mathbf{x})$, where $\mathbf{K}$ is the $n \times n$ matrix with $\mathbf{K}_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$, and $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n]^\top$.

The kernel method is related to MLPs through a kernel function called Neural Tangent Kernel [8]. For an MLP Π_θ with trainable parameters θ, it has been shown that, under certain conditions, the model Π_θ trained with stochastic gradient descent will converge to the estimate generated by the kernel method through Neural Tangent Kernel, which is defined as $\kappa_{\text{NTK}}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{E}_{\theta\sim\mathcal{N}} \left\langle \frac{\partial \Pi_\theta(\mathbf{x}_i)}{\partial \theta}, \frac{\partial \Pi_\theta(\mathbf{x}_j)}{\partial \theta} \right\rangle$. Importantly, for $\mathbf{x}_i,\ \mathbf{x}_j$ on the unit sphere, NTK is an inner product kernel, *i.e.*,

$$\kappa_{\text{NTK}}(\mathbf{x}_i, \mathbf{x}_j) = k_{\text{NTK}}(\mathbf{x}_i \cdot \mathbf{x}_j), \tag{5}$$

for some function k_{NTK}.

With the kernel function explained, we now show how positional encoding helps with expressiveness. By construction, the positional encoding maps pixel coordinates into sinusoidal vectors. This is closely related to random Fourier features which are provably able to approximate the family of shift-invariant kernel functions [24]. Specifically, a kernel function κ is shift-invariant if for any $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathbb{R}^d$, it holds that $\kappa(\mathbf{x}_1, \mathbf{x}_2) = \kappa(\mathbf{x}_1 - \mathbf{x}_2)$ with a slight abuse of notation. In other words, the value of $\kappa(\mathbf{x}_1, \mathbf{x}_2)$ only depends on the difference $\mathbf{x}_1 - \mathbf{x}_2$. It is clear that shift-invariance is an ideal property in imaging tasks.

To understand how the Fourier features generated by positional encoding can approximate shift-invariant kernel, we rewrite positional encoding (Eqn. 3) as:

$$\psi(\mathbf{x}) = [\sin(2\pi\mathbf{w}_1 \cdot \mathbf{x}), \cdots,\ \sin(2\pi\mathbf{w}_L \cdot \mathbf{x}), \cos(2\pi\mathbf{w}_1 \cdot \mathbf{x}), \cdots,\ \cos(2\pi\mathbf{w}_L \cdot \mathbf{x})],$$

where $\mathbf{x} = (\tilde{x}, \tilde{y})$, and $\mathbf{w}_i = [w_i, v_i]$ for $i = 1, \cdots, L$. Given any shift-invariant kernel $\kappa(\mathbf{x}_1, \mathbf{x}_2) = \kappa(\mathbf{x}_1 - \mathbf{x}_2)$, we define a distribution $\mathcal{P}$ over $\mathbf{w}$ as $\mathcal{P}(\mathbf{w}) = \frac{1}{2\pi}\int e^{-2\pi i \mathbf{w}^\top \mathbf{x}} \kappa(\mathbf{x}) d\mathbf{x}$, which is the Fourier transform of the kernel κ. Suppose that $\mathbf{w}_1, \cdots, \mathbf{w}_L$ are *i.i.d.* samples from $\mathcal{P}$. Then it holds that [24]:

$$\Pr\left[\sup_{\mathbf{x}_1,\mathbf{x}_2} \left| \frac{1}{L}\psi(\mathbf{x}_1)^\top\psi(\mathbf{x}_2) - \kappa(\mathbf{x}_1, \mathbf{x}_2)\right| \geq \epsilon\right] \leq 2^8 \frac{\sigma_{\mathcal{P}}^2}{\epsilon^2} \exp\left(-\frac{L\epsilon^2}{16}\right),$$

where $\sigma_{\mathcal{P}}^2 = \mathbb{E}_{\mathcal{P}}(\mathbf{w}^\top\mathbf{w})$. The above result indicates that with high probability, any shift-invariant kernel function can be approximated with Fourier features. Therefore, this demonstrates the expressive power of INR features.

Finally, we show that $\psi(\cdot)$ combined with the MLP Π_θ forms a shift-invariant kernel. We define the positional encoding kernel κ_{pe} as $\kappa_{\text{pe}}(\mathbf{x}_1, \mathbf{x}_2) = \psi(\mathbf{x}_1)^\top\psi(\mathbf{x}_2)$. It can be shown that $\kappa_{\text{pe}}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^L \cos(2\pi\mathbf{w}_j \cdot (\mathbf{x}_1 - \mathbf{x}_2))$ [28], which is shift-invariant. Combining with Eqn. 5, our positional encoding combined followed by MLP approximately equals to $k_{\text{NTK}}(\kappa_{\text{pe}}(\mathbf{x}_1 - \mathbf{x}_2))$, which is a shift-invariant kernel. This demonstrates the expressiveness of MORSE.