

Register Anything: Estimating “Corresponding Prompts” for Segment Anything Model

Shiqi Huang^{1,2}, Tingfa Xu², Wen Yan¹, Dean Barratt¹, and Yipeng Hu¹

¹ University College London, London, UK

² Beijing Institute of Technology, Beijing, China
{huangsq, ciom_xtf1}@bit.edu.cn
yipeng.hu@ucl.ac.uk

Abstract. Establishing pixel/voxel-level or region-level correspondences is the core challenge in image registration. The latter, also known as region-based correspondence representation, leverages paired regions of interest (ROIs) to enable regional matching while preserving fine-grained capability at pixel/voxel level. Traditionally, this representation is implemented via two steps: segmenting ROIs in each image then matching them between the two images. In this paper, we simplify this into one step by directly “searching for corresponding prompts”, using extensively pre-trained segmentation models (*e.g.*, SAM) for a training-free registration approach, PromptReg. Firstly, we introduce the “corresponding prompt problem”, which aims to identify a corresponding Prompt Y in Image Y for any given visual Prompt X in Image X, such that the two respectively prompt-conditioned segmentations are a pair of corresponding ROIs from the two images. Secondly, we present an “inverse prompt” solution that generates primary and optionally auxiliary prompts, inverting Prompt X into the prompt space of Image Y. Thirdly, we propose a novel registration algorithm that identifies multiple paired corresponding ROIs by marginalizing the inverted Prompt X across both prompt and spatial dimensions. Comprehensive experiments are conducted on five applications of registering 3D prostate MR, 3D abdomen MR, 3D lung CT, 2D histopathology and, as a non-medical example, 2D aerial images. Based on metrics including Dice and target registration errors on anatomical structures, the proposed registration outperforms both intensity-based iterative algorithms and learning-based DDF-predicting networks, even yielding competitive performance with weakly-supervised approaches that require fully-segmented training data.

Keywords: Image registration · Corresponding representation · Prompt engineering · Segment Anything Model (SAM).

1 Introduction

Image registration aligns two or more images to a common coordinate system, allowing for the overlay and analysis of corresponding features [7,10,26]. This

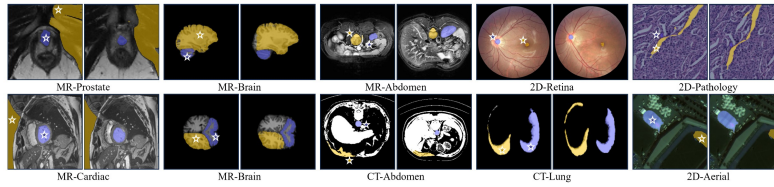


Fig. 1: PromptReg performs generic registration tasks with ROI-based correspondence representation. Given any point prompt \star , it identifies corresponding ROIs in paired images, where the blue and yellow indicate prompt locations inside and outside anatomical ROIs, respectively.

process is equivalent to establishing a ubiquitous correspondence at pixel/voxel level, that is, at each pixel/voxel location, a corresponding spatial location on another image is determined by registration. Pursuing such ubiquitous correspondence under this premise, the correspondence is often formulated by parametric spatial transformation [27], such as rigid and deformable models, or dense samples of displacement vectors at all discrete pixel/voxel locations, *i.e.*, dense displacement fields (DDFs) [8,2].

Between medical images, specific meanings of the correspondence may vary, for example, to indicate the same anatomy scanned at different time points or structures that share homology from different subjects. As pointed out in previous works [14,17], correspondence cannot (and need not) be defined necessarily everywhere (*i.e.*, at every pixel/voxel) in many clinical applications. Consequently, a new ROI-based representation has recently been proposed. This correspondence representation led to a new registration algorithm, which utilised the Segment Anything Model (SAM) to segment multiple regions-of-interest (ROIs) in two images and matched them in feature space [17], namely SAMReg.

SAMReg relied on the SAM’s unconditional segmentation capability to generate many candidate ROIs before matching their prototype features. However, this algorithm is inherently limited by its general-purposed formulation. First, the generated candidate ROIs cannot be controlled with respect to specific applications of interest. For instance, when prior knowledge on which class of ROIs should be considered more important, there is no mechanism to incorporate it. Second, the prototype matching is based on similarity (*e.g.* cosine similarity) between features of the candidate ROIs. This may be effective for a specific application, using carefully tuned hyperparameters such as the prototype dimensionality and the similarity threshold, but does not guarantee that the matched ROIs are of the same class between the two images.

In this study, we first develop methodologies for searching corresponding ROI pairs of the same class. Unlike the previously proposed unconditioned algorithm [17], the proposed approach exploits the prompt-conditioned SAM for, arguably, more controllable ROI class definitions. These classes may be i) pre-defined by prior clinical requirements to aid application-specific registration algorithms or facilitate an interactive registration; but can also be ii) randomly sampled for general-purpose, no-prior image registration algorithms. Focusing

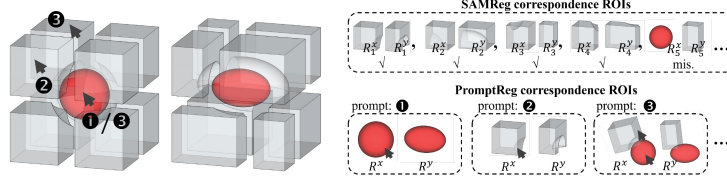


Fig. 2: Comparison of correspondence ROIs from different paradigms, where the red denotes the targeted ROI and the arrow represents the given prompts.

on point-based prompts, we formalize this as a “corresponding prompt problem” in Sec. 2. We then propose a solution by inverting a given prompt in Sec. 3.1.

With prompt-based segmentation, the inherent limitation in SAM and its generalisation to medical images became evident. For example, sensitivity in prompt localisation leads to unstable segmented ROI was reported [12,19], which was found more predominant in a domain-shifted dataset, such as medical images, and needed to resort a nontrivial fine-tuning effort [22,30]. For our registration algorithms, we further propose a marginalization strategy, over both the prompt sampling space and the transformation space, in Sec. 3.2.

In summary, our main contributions are: **1)** We propose a new image registration paradigm, which seeks correspondence-conditioned ROI pairs from two respective images. To our knowledge, it is the first promptable registration trial. **2)** Using SAM without any adaptation or fine-tuning, we introduce the general-purpose PromptReg by segmenting corresponding ROIs from a pair of images, based on any (manual or randomly-sampled) given vision prompts. **3)** Extensive experiments show that the proposed registration algorithms outperform the commonly-adopted iterative registration and unsupervised registration, and is competitive with weakly-supervised registration across five clinical applications.

2 Corresponding Prompt Problem

Revisiting ROI-based Correspondence Representation. The ROI-based correspondence representation for registration [17], spatial locations \mathbf{X} and \mathbf{Y} within a moving and a fixed image, I^x and I^y , respectively, is considered corresponding ROIs, represented through K pairs of ROI $\{(R_k^x, R_k^y)\}_{k=1}^K$. Here, $R_k^x = \{\mathbf{x}_l\}_{l=1}^{L_k^x}$ and $R_k^y = \{\mathbf{y}_l\}_{l=1}^{L_k^y}$ are two sets of L_k^x and L_k^y spatial discrete locations (*e.g.*, pixels/voxels) in the respective moving and fixed image spaces, $\mathbf{x}_l \in \mathbf{X}$ and $\mathbf{y}_l \in \mathbf{Y}$. This representation is thought necessary and effective for clinical practice. However, a dense correspondence representation, such as the DDF \mathcal{T} , can also be obtained by refining the region-specific alignment measure $\sum_{k=1}^K \mathcal{L}_{roi}(R_k^y, \mathcal{T}(R_k^x, \Theta))$, where Θ is parameters of the transformation \mathcal{T} .

The corresponding ROIs can be acquired through two ways: (1) *segmenting known shared classes*, or (2) *segmenting unknown classes and then matching them*. Currently, the former is conducted by weakly-supervised registration methods [15,14], which train the segmentor with predefined K -class groundtruth to predict intuitive corresponding ROIs of the same classes $\mathbf{C} = [C_1, \dots, C_K]^\top$,

$\mathbf{C} = \mathbf{C}^x = \mathbf{C}^y$, albeit with limited number of representation ROIs and narrow generalizability. The latter paradigm is implemented by recent SAMReg [17], consisting two independent segmentation without conditioning prompts, producing finite K^x and K^y classes \mathbf{C}^x and \mathbf{C}^y , with the highest class probabilities, respectively. Second, the subsequent *matching* identifies K shared classes \mathbf{C} , and $\{C_k\}_{k=1}^K = \{C_k^x\}_{k=1}^{K^x} \cap \{C_k^y\}_{k=1}^{K^y}$. In practice, it is this two-step approximation that limits the SAMReg to represent a required ROI correspondence that are locally accurate and densely sampled.

Why “Searching Corresponding Prompts” Instead of “Matching Corresponding ROIs”? A mirrored paradigm of SAMReg for *segmenting known shared classes at the same time* is to estimate two vision prompts Z_k^x and Z_k^y , for I^x and I^y , respectively, such that the common class C_k can be segmented by conditioning two SAM inferences using these two, namely, corresponding prompts. Compared with predefined classes in weakly-supervised methods, the class C_k , conditioned on prompt Z_k^x , requires neither specific labels nor anatomical knowledge, while still permitting targeted human interaction via Z_k^x .

The “Searching Corresponding Prompts” pattern is more cost-effective than SAMReg’s “Matching Corresponding ROIs,” both globally and locally. SAMReg cannot ensure that segmented classes from two independent SAM inferences intersect, $\{C_k^x\}_{k=1}^{K^x} \cap \{C_k^y\}_{k=1}^{K^y} \neq \emptyset$, nor guarantee a specific local ROI class C^* exists in both, $C^* \in (\{C_k^x\}_{k=1}^{K^x} \cap \{C_k^y\}_{k=1}^{K^y})$. In contrast, corresponding prompts resolve empty intersections by leveraging predefined (manual or automatic) target classes C^* (Fig. 2). For dense correspondence, SAMReg lacks control over ROI sampling. Even with many candidate ROIs, there is no guarantee that they or the matched set \mathbf{C} will align with the targeted ROIs. The finite sizes of \mathbf{C}^x and \mathbf{C}^y further constrain \mathbf{C} , increasing mismatches as more are assigned to \mathbf{C} .

What is the Definition of “Corresponding Prompt Problem”? Given images I^x and I^y , the goal is to estimate same-class ROIs. A vision prompt for I^x is $Z_k^x = \{\mathbf{x}_p\}_{p=1}^{P_k^x}$, where $\mathbf{x}_p \in \mathbf{X}$ and P_k^x is the number of discrete pixels/voxels. The pre-trained SAM segments ROI as $R_k^x = f^{SAM}(I^x, Z_k^x)$, assuming R_k^x belongs to class C_k .

The aim is to search for the “corresponding prompt” $Z_k^y = \{\mathbf{y}_p\}_{p=1}^{P_k^y}$, defined by P_k^y locations $\mathbf{y}_p \in \mathbf{Y}$ in the second image I^y , such that the segmented ROI $R_k^y = f^{SAM}(I^y, Z_k^y)$ in image I^y mirrors the same class C_k of the ROI R_k^x in image I^x and (R_k^x, R_k^y) is considered as the “corresponding ROI”. This paradigm allows for the investigation of class-consistency across different images, leveraging the capabilities of the SAM or its variants to perform corresponding ROI segmentation, based on either predefined or arbitrary visual prompts.

3 PromptReg: Promptable Registration Algorithm via Inverse Prompt Engineering

The proposed PromptReg comprises two steps: (inverted) prompt searching and prompt marginalization. In the first step, prompt searching infers the correspond-

ing prompt on the target image, while in the second step, prompt marginalization leverages an augmentation-based strategy to improve robustness.

3.1 Prompt Searching

Prompt searching aims to map a given prompt Z_k^x from an image I^x to a corresponding prompt set Z_k^y on the other image I^y , ensuring both represent the same class C_k . In this section, the prompt set Z_k^y is set to include a primary prompt, derived through an inversion process linking C_k to I^y , and an optional auxiliary prompt. When the primary prompt alone cannot effectively represent C_k , (e.g., due to large image differences or low prominence of C_k), the auxiliary prompt, located at the foreground-background boundary, supplements Z_k^y .

Primary Prompt. In a typical SAM inference for segmenting ROI R_k^x from image I^x , an image encoder \mathcal{E}^{im} embeds $I^x \in \mathbb{R}^{H \times W \times D}$ into features $F^x \in \mathbb{R}^{H' \times W' \times D' \times N'}$. A prompt encoder \mathcal{E}^{pr} embeds the prompt Z_k^x into a class encoding C_k , which may be unknown or unspecified, determined by Z_k^x . The mask decoder \mathcal{D} then decodes F^x , conditioned on C_k , to produce the segmented ROI R_k^x . R_k^x is further abstracted as a prototype $G_k = f^{proto.}(R_k^x, F^x)$, where $f^{proto.}$ is the prototype extraction function iterating all voxel locations, commonly used in prototypical few-shot learning [9,16,5,31]. The resulting prototype $G_k \in \mathbb{R}^{1 \times N'}$ serves as an alternative representation of the class C_k .

G_k is considered a function of only one variable Z_k^y , i.e., $G_k = f(Z_k^y) = f^{proto.}(\mathcal{D}(F^x, \mathcal{E}^{pr}(Z_k^y)), F^x)$. Using the first-order Taylor expansion $G_k \approx f^0 + \mathcal{J} \cdot Z_k^y$ obtains a general solution for $Z_k^y \propto \mathcal{J}^{-1} \cdot G_k$, where f^0 is a constant and $\mathcal{J} \in \mathbb{R}^{N' \times 3}$ (given 1 point prompt) is the Jacobian with respect to the function f , specifically, $\mathcal{J} = \frac{\partial f}{\partial Z_k^y} = \frac{\partial f^{proto.}}{\partial \mathcal{D}} \cdot \frac{\partial \mathcal{D}}{\partial \mathcal{E}^{pr}} \cdot \frac{\partial \mathcal{E}^{pr}}{\partial Z_k^y}$.

To inverse the inference with image I^y and unknown prompt Z_k^y , the logits of class probability maps $S_k^{tx}, S_k^{ty} \in \mathbb{R}^{H' \times W' \times D'}$ are generated using cosine similarity function f^{sim} , e.g., $S_k^{tx} = f^{sim}(G_k, F^x)$, where $S_{k(i')}^{tx} = \frac{G_k \cdot F_{(i')}^x}{\|G_k\| \cdot \|F_{(i')}^x\|}$, where $i' \in \mathbb{R}^3$ are all the 3D “spatial” locations in the feature space. The bracket subscripts (i') is denoted to index the scalars and vectors, from S_k^{tx} and F^x , respectively. With the same G_k , $S_k^{ty} = f^{sim}(G_k, F^y)$ is computed.

Following the Taylor expansion, $S_k^{ty} \approx (f^{sim})^0 + \mathcal{J}^{sim} \cdot G_k$, leads to the association between S_k^{ty} and Z_k^y that

$$Z_k^y = (\mathcal{J}^{sim})^{-1} \cdot \mathcal{J}^{-1} \cdot S_k^{ty} + \rho, \quad (1)$$

where ρ is a constant set manually and the similarity Jacobian matrix $\mathcal{J}^{sim} \in \mathbb{R}^{H' \times W' \times D' \times N'}$ is derived by $\mathcal{J}^{sim} = \frac{\partial S_k^{ty}}{\partial G_k}$, where $\frac{\partial S_{k(i')}^{ty}}{\partial G_k} = \frac{F_{(i')}^y - S_{k(i')}^{ty} \cdot \frac{G_k}{\|G_k\|}}{\|F_{(i')}^y\| \cdot \|G_k\|}$.

The obtained inverse prompt Z_k^y corresponds to the given prompt Z_k^x with the same structural format, i.e., it adheres to a one-to-one or multiple-to-multiple correspondence between points.

Auxiliary Prompts. Conditioned on the primary prompt Z_k^y , the segmented ROI $R_k^y = \mathcal{D}(F^y, \mathcal{E}^{pr}(Z_k^y))$ is generated. Ideally, R_k^y should closely match the corresponding ROI, R_k^x , produced from prompt Z_k^x . However, R_k^x and R_k^y may not

align in some cases, for example, R_k^y partially capture a complete structure identified by R_k^x . To address these mismatches, more precise prompts are necessary to define the desired ROI R_k^y .

Given the inherent consistency of registration tasks, misaligned ROIs (R_k^x and R_k^y) exhibit abnormal shapes. To quantify this, we compute the Hausdorff distance $d_H^{x \rightarrow y}$ and $d_H^{y \rightarrow x}$ between their contours, identifying points of maximum discrepancy \mathbf{x}^* and \mathbf{y}^* . If $\|d_H^{x \rightarrow y} - d_H^{y \rightarrow x}\| > \sigma$, where σ is empirically determined, \mathbf{x}^* or \mathbf{y}^* is considered an anomaly.

The prompt is iteratively placed until convergence, *i.e.*, $\|d_H^{x \rightarrow y} - d_H^{y \rightarrow x}\| < \sigma$. In s -th iteration, based on the identified anomaly, we generate a new prompt, $\mathbf{z}_{k,s}^y$. This prompt can be either a negative point or a positive point:

$$\mathbf{z}_{k,s}^y = \begin{cases} \mathbf{y}^* \text{ (neg)}, & d_H^{x \rightarrow y} > d_H^{y \rightarrow x}, \\ \mathbf{x}^* + \epsilon \nabla f^{x \rightarrow y}(\mathbf{r})|_{\mathbf{r}=\mathbf{x}^*} \text{ (pos)}, & d_H^{x \rightarrow y} < d_H^{y \rightarrow x}, \end{cases} \quad (2)$$

where $f^{x \rightarrow y}(\mathbf{r}) = \frac{1}{2} \|\mathbf{r} - \mathbf{y}^*\|^2$. These prompts are iteratively incorporated into Z_k^y to progressively align R_k^y with the R_k^x .

3.2 Prompt Marginalization

Probabilistic Notes. The SAM models class probability as $\mathcal{R}_k = p(C_k | \mathbf{I}, \mathbf{Z}_k)$. Marginalizing over \mathbf{Z}_k and $p(C_k | \mathbf{I}) = \int_{\mathbf{Z}_k \in \Omega_{Z_k^x}} p(C_k | \mathbf{I}, \mathbf{Z}_k) p(\mathbf{Z}_k | \mathbf{I}) d\mathbf{Z}_k$, where $\Omega_{Z_k^x}$ defines the prompt sampling space on I^x . Multiple prompts enable marginalization, capturing \mathbf{Z}_k uncertainty and improving spatial coverage (*i.e.*, $K > 1$).

For a second image I^y , estimating $\mathbf{Z}_k = Z_k^y$ considers it a parameter linked to the prototype G_k^x from Z_k^x (Sec. 3.1). The marginal likelihood with fixed I^y is: $\int_{\mathbf{R}_k^y \in \Omega_{R_k^y}} p(\mathbf{R}_k^y | Z_k^y, I^y) d\mathbf{R}_k^y = \int_{\mathbf{G}_k^x \in \Omega_{G_k^x}} p(\mathbf{G}_k^x | Z_k^y, I^y) d\mathbf{G}_k^x = \int_{\mathbf{Z}_k \in \Omega_{Z_k^x}, \mathbf{I} \in \Omega_{I^x}} p(\mathbf{Z}_k, \mathbf{I} | Z_k^y, I^y) d(\mathbf{Z}_k, \mathbf{I})$. Here, \mathbf{R}_k^y and \mathbf{G}_k^x denote the random variables of segmented ROI and prototype, and Ω_* defines the sampling space. Estimating this marginal likelihood follows Bayesian model averaging [25].

Spatial Transformation for Aggregation. To sample \mathbf{Z}_k and \mathbf{I} from respective $\Omega_{Z_k^x}$ and Ω_{I^x} , one could argue the benefit in applying independent spatial transformation \mathcal{A}_j to each. However, to ensure the transformed prompt $\tilde{Z}_{k,j}^x = Z_k^x \circ \mathcal{A}_j$ and image $\tilde{I}_j^x = I^x \circ \mathcal{A}_j$ represent the same class C_k , the same spatial transformation \mathcal{A}_j is applied.

Based on the J transformed images $\{\tilde{I}_j^x\}$ and prompts $\{\tilde{Z}_{k,j}^x\}$, a set of ROIs (class probability maps) $\{\tilde{\mathcal{R}}_{k,j}^x\}$ are predicted. Using the inverse-transformed ROIs $\{\tilde{\mathcal{R}}_{k,j}^x \circ \mathcal{A}_j^{-1}\}$, a set of corresponding prompts are computed $\{\tilde{Z}_{k,j}^y\}$, which obtains a set of corresponding ROIs $\{\tilde{\mathcal{R}}_{k,j}^y\}$, for each class C_k . The final pair of corresponding ROIs $(\bar{\mathcal{R}}_k^x, \bar{\mathcal{R}}_k^y)$ are averages: $\bar{\mathcal{R}}_k^x = \frac{1}{J} \sum_{j=1}^J \tilde{\mathcal{R}}_{k,j}^x$, $\bar{\mathcal{R}}_k^y = \frac{1}{J} \sum_{j=1}^J \tilde{\mathcal{R}}_{k,j}^y$.

4 Experiments and Results

Datasets and Evaluation Metrics: We evaluate our method in five medical and non-medical datasets: MR-Prostate [1], MR-Abdomen [18], CT-Lung [13],

2D-Pathology [20] and 2D-Aerial [23]. The first three datasets are three-dimensional and contain MR and CT imaging modalities, while the latter two are two-dimensional. To ensure a fair comparison with other supervised registration methods, 80% of the images are utilized for training purposes. Our method does not use this subset for training and is tested on the remaining data alongside other methods. Additionally, for ablative experiments, the proposed algorithm is implemented to all available images. Regarding registration strategies, inter-subject registration is applied to the MR-Prostate and MR-Abdomen datasets to align images to a standard reference. Conversely, intra-subject registration applies for the lung, histological, and aerial images to align image pairs that acquired at varying time. Metrics include Dice and target registration error (TRE). For non-rigid registrations, Dice is assessed on moved and fixed critical anatomical structures, with TRE derived from the centroids of these ROIs.

Implementation Details: In our evaluation experiments, the prompt is randomly sampled four times to infer dense correspondence. Hyperparameters σ and ϵ for auxiliary prompt setting is set to 20.0 and 2.0, respectively. For marginalization in spatial transformation, transformations such as flipping, rotation, and scaling are randomly sampled, with flipping consisting of horizontal and vertical, rotation ranging from 0° to 360° , scaling ranging from 0.5 to 2.5. The implementation is based on PyTorch and MONAI [3]. The algorithm was executed on an NVIDIA Quadro GV100. A code demo is available at: [PromptReg](#).

Methods	MR-Prostate		MR-Abdomen		CT-Lung		2D-Pathology		2D-Aerial	
	Dice	TRE	Dice	TRE	Dice	TRE	Dice	TRE	Dice	TRE
NiftyReg[24]	7.68±3.98	4.67±3.48	8.93±2.21	3.13±2.89	10.93±2.02	4.23±1.64	6.81±3.02	5.90±3.75	10.21±3.02	4.02±2.25
VoxelMorph[2]	55.94±3.34	3.68±1.98	58.1±3.95	2.76±2.41	77.98±2.72	3.24±0.81	59.34±3.72	4.31±2.13	72.73±2.41	3.53±1.30
LabelReg[15]	76.72±3.23	2.72±1.23	75.97±2.42	1.56±1.34	83.56±2.43	1.52±0.86	-	-	-	-
KeyMorph[11]	70.52±3.25	3.19±1.63	72.77±2.56	2.21±1.87	81.71±2.26	2.93±0.86	73.15±2.39	3.47±2.01	80.22±2.13	3.05±1.19
TransMorph[4]	71.77±3.01	3.11±1.27	73.81±2.52	2.10±1.76	81.97±2.19	2.84±0.79	75.42±2.08	3.29±1.95	82.63±2.09	2.91±1.02
SAMReg[17]	75.67±3.19	2.09±1.22	73.65±2.52	1.43±1.21	85.23±2.16	1.31±0.91	69.87±3.70	3.12±1.23	85.34±2.31	2.57±1.10
PromptReg w SAM[21]	76.48±2.92	2.10±2.68	72.48±3.57	2.38±1.13	87.76±3.21	2.02±1.03	72.71±3.63	3.41±0.91	86.65±1.80	2.62±1.01
PromptReg w MedSAM[22]	72.91±3.12	3.32±2.30	73.86±3.37	2.18±1.22	88.40±3.23	1.66±0.99	71.47±3.52	3.14±1.34	83.77±2.03	3.69±1.78
PromptReg w SAMed2d[6]	74.28±3.39	3.70±2.04	74.55±3.54	1.86±1.05	88.03±3.10	2.06±0.84	87.35±3.89	3.90±1.97	82.69±2.13	2.86±1.65
PromptReg w SAMed3d[29]	75.63±2.61	3.95±1.59	75.35±2.99	1.72±1.01	88.44±2.83	1.73±0.84	-	-	-	-
PromptReg w 3dAdapter[30]	77.76±2.20	2.06±0.91	77.3±2.53	1.22±0.92	90.46±2.53	1.20±0.82	-	-	-	-

Table 1: Comparison of different registration models and our PromptReg with various promptable models across five datasets, highlighting the best 3D model results in **bold** and the best 2D model results with an underline.

Comparison with SoTA Methods. Table 1 presents a quantitative comparison across various medical image registration methods, including NiftyReg [24] that represents a class of non-learning iterative registration algorithms [28], SAMReg [17], an existing SAM-based training-free algorithm, and also contrast them, albeit less equitably, unsupervised methods VoxelMorph, KeyMorph and latest TransMorph [2], and ROI-supervised LabelReg [15], on five datasets. Morph variants and LabelReg have been proposed for medical image research datasets, with LabelReg additionally requiring fully-segmented anatomical annotations. As shown in Table 1, PromptReg consistently outperforms these methods across all datasets without dataset-specific tuning, particularly competitive in intra-subject registration tasks perhaps due to higher prevalence in consistent intra-subject ROIs. Compared with SAMReg (5 paired ROIs), PromptReg

Prompt Location		Dice	TRE
outsideROI	insideROI		
✓	✓	77.76±2.20	2.06±0.91
✓		75.23±3.11	2.83±1.06
	✓	81.77±3.08	1.90±0.91

Table 2: Ablation of given prompt Z_k^x location on MR-Prostate dataset, where *ROI* denotes the target anatomical region.

# of Prompts	Dice	TRE
1	70.07±3.61	4.07±2.03
2	72.87±2.44	3.42±1.38
3	75.32±2.31	2.64±1.42
4	76.72±2.33	2.27±1.29
5	77.34±2.37	2.06±1.25
6	77.84±2.32	2.00±1.21

Table 4: Ablation of given prompt Z_k^x quantity on MR-Prostate dataset.

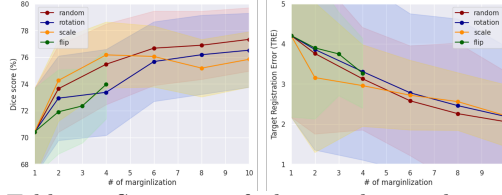


Table 5: Statistics of the number and type of marginalization iterations on MR-Prostate dataset.

demonstrates lower variance that may indicate greater robustness. Additionally, when local and/or dense correspondence is required clinically, the PromptReg exhibited superior capability to generate more paired ROIs that are indeed corresponding to each other - a limitation of SAMReg discussed in Sec. 2.

Comparison of Promptable Models. Promptable models can be directly integrated into the PromptReg paradigm, facilitating the adoption of cutting-edge segmentation models for registration. Table 1 evaluates various models with four random prompts each. For 3D images processed with 2D models (SAM, MedSAM, SAMed2d), slices are handled individually. Results show that 3D-specific models (SAMed3d, 3DAdapter) generally perform better due to their 3D awareness. SAM, however, excels in non-medical tasks due to its generalizability. MedSAM and SAMed2d perform well on the Abdomen dataset, benefiting from pre-training on abdominal data. However, prior exposure doesn’t always improve registration, *e.g.*, SAMed2d, pre-trained on prostate data, tends to focus on central structures, leading to inconsistencies when prompted outside these ROIs.

Ablation on Given Prompt Z_k^x . **a)** Location. PromptReg allows predefined prompt locations. Constraining prompts within ROIs significantly improves registration, especially in inter-subject cases (Table 2). **b)** Quantity. More prompts enhance performance but with diminishing returns, improving ROI alignment and downstream tasks (Table 4).

Ablation on Inverse Prompt Z_k^y . **a)** Threshold σ . Table 3 shows that $\sigma = 20.0$ optimizes registration by balancing mismatch tolerance and auxiliary prompts. **b)** Transformation \mathcal{A} . Fig. 5 shows that scaling outperforms rotation for < 7 iterations, while a random strategy works best for more.

5 Conclusion

In this study, we introduce a new “corresponding prompt problem” that reformulates the image registration to search corresponding prompts for pre-trained segmentation models, with a proposed promptable registration algorithm, PromptReg; comprehensive experiments highlight its competitive performance, indicating a new direction for registration research and a new application for vision foundation models.

Acknowledgments. This work was supported by the International Alliance for Cancer Early Detection, an alliance between Cancer Research UK [C28070/A30912; C73666/A31378], Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer Institute, University College London and the University of Manchester. This work was also supported by the National Institute for Health Research (NIHR) University College London Hospitals (UCLH) Biomedical Research Centre (BRC).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmed, H.U., Bosaily, A.E.S., Brown, L.C., Gabe, R., Kaplan, R., Parmar, M.K., Collaco-Moraes, Y., Ward, K., Hindley, R.G., Freeman, A., et al.: Diagnostic accuracy of multi-parametric mri and trus biopsy in prostate cancer (promis): a paired validating confirmatory study. *The Lancet* **389**(10071), 815–822 (2017)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging* **38**(8), 1788–1800 (2019)
3. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701* (2022)
4. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis* **82**, 102615 (2022)
5. Chen, S., Xie, M.: Augmenting advertiser decision support with generative ai and interactive analytics. *Authorea Preprints* (2025)
6. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. *arXiv preprint arXiv:2308.16184* (2023)
7. De Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis* **52**, 128–143 (2019)
8. DeTone, D., Malisiewicz, T., Rabinovich, A.: Deep image homography estimation. *arXiv preprint arXiv:1606.03798* (2016)
9. Dong, N., Xing, E.P.: Few-shot semantic segmentation with prototype learning. In: *BMVC*. vol. 3, p. 4 (2018)
10. Eppenhof, K.A., Lafarge, M.W., Moeskops, P., Veta, M., Pluim, J.P.: Deformable image registration using convolutional neural networks. In: *Medical Imaging 2018: Image Processing*. vol. 10574, pp. 192–197. SPIE (2018)

11. Evan, M.Y., Wang, A.Q., Dalca, A.V., Sabuncu, M.R.: Keymorph: Robust multi-modal affine registration via unsupervised keypoint detection. In: MIDL (2022)
12. He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., Guo, Z., Li, X.: Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. NeurIPS (2024)
13. Hering, A., Murphy, K., van Ginneken, Bram: Learn2reg challenge: Ct lung registration. <https://doi.org/10.5281/zenodo.3835682> (2020)
14. Hu, Y., Gibson, E., Barratt, D.C., Emberton, M., Noble, J.A., Vercauteren, T.: Conditional segmentation in lieu of image registration. In: MICCAI (2019)
15. Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C.M., Emberton, M., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. *Medical image analysis* **49**, 1–13 (2018)
16. Huang, S., Xu, T., Shen, N., Mu, F., Li, J.: Rethinking few-shot medical segmentation: a vector quantization view. In: CVPR (2023)
17. Huang, S., Xu, T., Shen, Z., Saeed, S.U., Yan, W., Barratt, D., Hu, Y.: One registration is worth two segmentations. In: MICCAI (2024)
18. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. *Medical Image Analysis* **69**, 101950 (2021)
19. Ke, L., Ye, M., Danelljan, M., Tai, Y.W., Tang, C.K., Yu, F., et al.: Segment anything in high quality. NeurIPS (2024)
20. Kevin, E., Bin, L., Adib, K.: Multimodal biomedical dataset for evaluating registration methods (full-size tma cores). <https://doi.org/10.5281/zenodo.4550300> (2021)
21. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
22. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. *Nature Communications* **15**(1), 654 (2024)
23. Michele, V., Vittorio, F.: Zurich summer dataset. <https://doi.org/10.5281/zenodo.5914759> (2022)
24. Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S.: Global image registration using a symmetric block-matching approach. *Journal of medical imaging* **1**(2), 024003–024003 (2014)
25. Raftery, A.E., Madigan, D., Hoeting, J.A.: Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92**(437), 179–191 (1997)
26. Rohé, M.M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: Svf-net: learning deformable image registration using shape matching. In: MICCAI (2017)
27. Rohr, K., Stiehl, H.S., Sprengel, R., Buzug, T.M., Weese, J., Kuhn, M.: Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on medical imaging* **20**(6), 526–534 (2001)
28. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging* **18**(8), 712–721 (1999). <https://doi.org/10.1109/42.796284>
29. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., et al.: Sam-med3d. arXiv preprint arXiv:2310.15161 (2023)

30. Wu, J., Fu, R., Fang, H., Liu, Y., Wang, Z., Xu, Y., Jin, Y., Arbel, T.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
31. Yang, H.M., Zhang, X.Y., Yin, F., Liu, C.L.: Robust classification with convolutional prototype learning. In: CVPR (2018)