ViPS-FSL: Virtual-Patient Densification via Latent Space Co-Translation for Robust Few-Shot Cardiac MRI Segmentation

Mohamed Elbayumi[1,2] and Mohammed S. M. Elbaz[1*]

[1] Department of Radiology, Northwestern University, Chicago, IL, USA
[2] Department of Biomedical Engineering, Northwestern University, Chicago, IL, USA
mohammed.elbaz@northwestern.edu

Abstract. Few-shot cardiac MRI segmentation is often limited by support-set underutilization and reliance on unlabeled data, degrading performance on unseen domains and compromising the functional measurements (e.g., ejection fraction, chamber volumes) central to cardiac diagnosis. We propose ViPS-FSL, which reformulates few-shot segmentation as support-set densification: from a sparse, pathology-diverse support set, it builds a cross-pathology pairwise support graph with phase-aware edge typing to learn local image–mask co-translation models, then samples a graph-induced Perceptual Feature Space via volume-calibrated latent interpolation to decode fully labeled virtual patients. A segmentation model is then trained on the labeled support set augmented with these virtual patients, using no unlabeled data. We evaluate in two settings: in-distribution on a held-out same-domain ACDC test split, and out-of-distribution on the more challenging M&Ms cohort, unseen and spanning multiple vendors, centers, and pathologies. On held-out ACDC (N=50), ViPS-FSL ranks first at both 7 and 11 support subjects (88.5/89.0 average Dice), outperforming the strongest baseline by +2.6/+1.4 Dice with no unlabeled data. Under single-source transfer to the unseen M&Ms cohort (N=320), it outperforms five state-of-the-art baselines—each using 63/59 unlabeled subjects—reaching 87.4%/3.59 mm and 88.1%/3.29 mm (Dice/HD95), with per-chamber gains up to 5.5 Dice and 39% lower HD95 and generalizing across three unseen vendors despite single-vendor supervision. Thus, densifying labeled support—rather than relying on unlabeled data—yields strong, domain-robust few-shot segmentation. Code: https://github.com/Elbaz-PIC-CVI-Lab/ViPS-FSL.

Keywords: Cardiac MRI Segmentation · Support-Set Densification · Cross-Domain Generalization · Few-Shot Segmentation.

1 Introduction

Deep learning has transformed cardiac MRI segmentation, which underpins the quantitative functional measurements (e.g. ejection fraction or myocardial mass) that drive cardiac diagnosis and treatment planning [1]. Achieving this

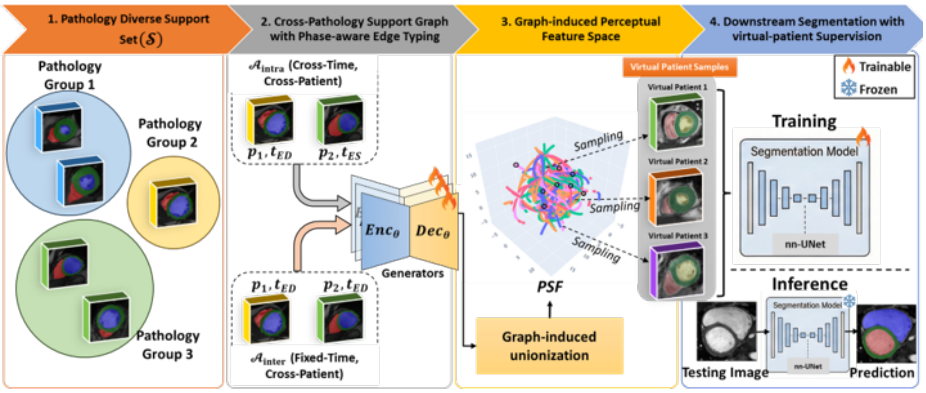

Fig. 1. Overview of the proposed ViPS-FSL framework.

reliably requires large, diverse annotated datasets that generalize across scanners, vendors, and populations [2, 3], yet expert annotation is costly. The problem is acute in cardiac MRI, where clinically meaningful variability is jointly shaped by pathology-dependent remodeling and cardiac phase [4], so sparse annotations rarely cover the relevant state space; this leads to segmentation performance degradation on unseen scanners or pathologies, and consequently inaccurate functional parameters that affect the diagnosis. Few-Shot Learning (FSL) has therefore emerged as a promising direction for low-label cardiac segmentation [5–7]. Existing FSL paradigms share recurring limitations: prototype/metric-based methods transfer class representations from a small support set [8–11], but struggle when sparse examples do not cover pathology-induced variation. Semi-supervised approaches improve low-label performance by leveraging unlabeled data [12–16], yet their gains depend on the source-domain unlabeled distribution and weaken under out-of-domain transfer. Recently, foundation-model distillation strategies show promise [17–19], but still require adaptation and do not model the coupled effects of remodeling and cardiac-phase dynamics. Collectively, these expose three bottlenecks: (B1) *support-set underutilization*, where sparse examples are treated as isolated templates; (B2) *unlabeled-data dependency*, where gains rely on unlabeled cohorts, introducing cohort-specific bias; and (B3) *cross-domain gap*, where models fail on unseen pathology–phase combinations in new datasets. These are amplified in cardiac MRI, where variability is jointly driven by cross-pathology remodeling and phase-dependent deformation.

Rather than relying on additional supervision, ViPS-FSL maximizes the utilization of the support set it already has: we decode labeled virtual patients from a learned feature space and treat them as augmentation samples that densify the labeled support set. This reframes few-shot segmentation as support-set densification—densification is the mechanism, improved single-source cross-domain segmentation the objective. From a pathology-diverse support set in which each phenotype contributes ED and ES states, ViPS-FSL builds a cross-pathology pairwise support graph with phase-aware edge typing—here "graph"

denotes pairwise support relations—and learns local image–mask co-translation models, one per edge×phase-pair. Their latent segments define a graph-induced Perceptual Feature Space (PFS), sampled via volume-calibrated interpolation to decode fully labeled virtual 3D image–mask pairs that augment the labeled support set, on which a segmentation network [2] is trained, using no unlabeled data.

Our contributions are: (1) a cross-pathology, phase-aware pairwise support graph with per-edge co-translation that converts sparse support into structured, labeled augmentation supervision; (2) a graph-induced PFS with volume-calibrated interpolation that densifies labeled support into a fully labeled virtual-patient cohort from support only; and (3) a strict two-setting evaluation — in-distribution (ACDC, N=50) where ViPS-FSL ranks first at two support-set sizes (7 and 11) using no unlabeled data, and single-source cross-domain transfer (ACDC [1]→M&Ms [23]) where it achieves the best segmentation with consistent gains across three unseen vendors.

2 Methods

2.1 Problem Formulation

Let $\mathcal{D} = \{s_i\}_{i=1}^{N}$ denote a cardiac MRI dataset of N subjects. Each subject:

$$s_i = \big(\{(x_i^t, y_i^t)\}_{t\in\mathcal{T}},\, g_i\big) \tag{1}$$

where $x_i^t \in \mathbb{R}^{H\times W\times Z}$ is a 3D cardiac MRI volume at cardiac phase t, $y_i^t \in \{0,1,\ldots,C\}^{H\times W\times Z}$ is the corresponding multi-class segmentation mask with C anatomical classes, and g_i denotes the pathology/phenotype label (e.g., healthy, hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), myocardial infarction (MI)). The cardiac phase set $\mathcal{T} = \{\mathrm{ED}, \mathrm{ES}\}$, with ED and ES as end-diastole and end-systole.

In the few-shot learning (FSL) setting, a small, labeled support set:

$$\mathcal{S} = \{s_i\}_{i=1}^{n} \subset \mathcal{D}, \quad n \ll N \tag{2}$$

$\mathcal{S}$ is selected for training. The key challenge is that $\mathcal{S}$ sparsely samples the cardiac state space, which is jointly shaped by inter-subject anatomy, pathology-related remodeling, and phase-dependent deformation. This results in bottlenecks (B1–B3) mentioned in the introduction. ViPS-FSL addresses these bottlenecks by reformulating few-shot supervision as support-set densification. In our setting, $\mathcal{S}$ is chosen to be pathology-diverse (at least one representative per phenotype). We model a support-conditioned translator $\mathcal{G}$ that maps sparse labeled support to a dense virtual labeled cohort:

$$\mathcal{V} = \mathcal{G}(\mathcal{S}) = \{(\tilde{x}_k, \tilde{y}_k)\}_{k=1}^{M} \tag{3}$$

where $\mathcal{V}$ is a fully labeled virtual cohort constructed strictly from $\mathcal{S}$, with $\tilde{x}_k \in \mathbb{R}^{H\times W\times Z}$ the k-th 3D volume and $\tilde{y}_k \in \{0,1,\ldots,C\}^{H\times W\times Z}$ its multi-class mask. $\mathcal{V}$ covers intermediate pathology–phase states absent from $\mathcal{S}$. A full overview of the method is given in Fig. 1.

2.2 Cross-Pathology Support Graph with Phase-Aware Edge Typing

To operationalize $\mathcal{V} = \mathcal{G}(\mathcal{S})$, ViPS-FSL first constructs a non-redundant cross-phenotype, pathology-diverse pairwise support graph over $\mathcal{S}$ as $\mathcal{H}(\mathcal{S}) = (\mathcal{S}, \mathcal{E})$. Graph nodes represent labeled support subjects, while edges encode pairwise cross-subject relations as local latent traversal paths. Phase-aware edge typing captures temporal dynamics, and cross-pathology coupling is established through pairings of distinct patient phenotypes. We define the set of unordered support-subject pairs as:

$$\mathcal{P}(\mathcal{S}) = \{\, (i, j) \mid 1 \leq i < j \leq n \,\} \tag{4}$$

which avoids redundant reverse instantiations (i.e., (i, j) and (j, i) are not treated as separate pair instances). Hence, $|\mathcal{P}(\mathcal{S})| = \binom{n}{2}$. For each pair $(i, j) \in \mathcal{P}(\mathcal{S})$, let $a, b \in \mathcal{T}$ denote source and target phases. We define two phase-pair edge-type sets:

$$\mathcal{A}_{\text{intra}} = \{\, (a, b) \in \mathcal{T} \times \mathcal{T} \mid a = b \,\} = \{(\text{ED}, \text{ED}), (\text{ES}, \text{ES})\} \tag{5}$$

$$\mathcal{A}_{\text{inter}} = \{\, (a, b) \in \mathcal{T} \times \mathcal{T} \mid a \neq b \,\} = \{(\text{ED}, \text{ES}), (\text{ES}, \text{ED})\} \tag{6}$$

While $\mathcal{A}_{\text{intra}}$ preserves cardiac phase, modeling cross-subject/cross-pathology variation, $\mathcal{A}_{\text{inter}}$ jointly models cross-subject/cross-pathology variation with phase-dependent deformation/remodeling. Thus, subject pairing controls cross-pathology traversal, while phase-aware edge typing controls how each pair is traversed in latent space.

2.3 Graph-Indexed Local Co-Translation Models

Rather than learning a single global translator, ViPS-FSL learns a set of local pairwise co-translation models indexed by graph edges:

$$\mathcal{F} = \{\, F_{ij}^{a \leftrightarrow b} \mid (i, j) \in \mathcal{P}(\mathcal{S}),\ (a, b) \in \mathcal{A}_{\text{intra}} \cup \mathcal{A}_{\text{inter}} \,\} \tag{7}$$

Each local bidirectional model $F_{ij}^{a \leftrightarrow b}$ is trained on labeled state pairs $(x_i^a, y_i^a) \leftrightarrow (x_j^b, y_j^b)$, enabling both forward and reverse traversals within a single model and yielding a piecewise latent representation of the support-induced cardiac state space. To preserve image–label coupling, translation is performed jointly in concatenated image–mask space; given $(x_{\text{src}}, y_{\text{src}})$ and $(x_{\text{tgt}}, y_{\text{tgt}})$, each local model uses a 3D CNN encoder–decoder generator that maps a source state to a translated image–mask pair $(x_{\text{gen}}, y_{\text{gen}}) = G_{\theta_{ij}}^{(a,b)}(x_{\text{src}}, y_{\text{src}})$. A parallel 3D Vision-Transformer registration network $R_{\psi_{ij}}^{(a,b)}$, adopted from [20], predicts a deformation field ϕ between the generated and the target samples:

$$\phi = R_{\psi_{ij}}^{(a,b)}\big((x_{\text{gen}}, y_{\text{gen}}),\ (x_{\text{tgt}}, y_{\text{tgt}})\big) \tag{8}$$

This co-translation pairs appearance and label changes, optimizing each model by:

$$\mathcal{L}_{\text{local}} = \lambda_{\text{reg}}\mathcal{L}_{\text{reg}} + \lambda_{\text{fid}}\mathcal{L}_{\text{fid}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}} \tag{9}$$

The registration loss enforces alignment with the target and smooth deformation:

$$\mathcal{L}_{\text{reg}} = \mathcal{L}_{\text{corr}}\big((x_{\text{gen}} \circ \phi,\, y_{\text{gen}} \circ \phi),\, (x_{\text{tgt}}, y_{\text{tgt}})\big) + \lambda_{\text{sm}} \|\nabla \phi\|_2^2 \tag{10}$$

where $\mathcal{L}_{\text{corr}}$ is an L_1-based correction term jointly applied to image and mask channels. To stabilize translation, the image fidelity loss combines Learned Perceptual Image Patch Similarity (LPIPS) [21] and Structural Similarity Index Measure (SSIM):

$$\mathcal{L}_{\text{fid}} = \lambda_{\text{LPIPS}}\, \text{LPIPS}(x_{\text{gen}}, x_{\text{tgt}}) + \lambda_{\text{SSIM}}\big(1 - \text{SSIM}(x_{\text{gen}}, x_{\text{tgt}})\big) \tag{11}$$

To preserve anatomical label structure, we use a class-weighted mask consistency loss:

$$\mathcal{L}_{\text{mask}} = \sum_{c=1}^{C} \omega_c \Big(\mathcal{L}_{\text{GDice}}^{(c)} + \lambda_{\text{mSSIM}} \mathcal{L}_{\text{SSIM}}^{(c)} + \lambda_{\text{NCC}} \mathcal{L}_{\text{NCC}}^{(c)} \Big) \tag{12}$$

with GDice as generalized Dice loss and NCC as normalized cross-correlation.

2.4 Perceptual Feature Space and Virtual Cohort Construction

We define the Perceptual Feature Space (PFS) as the union of latent segments induced by all local co-translation models over the cross-pathology support graph:

$$\mathcal{M}_{\text{PFS}} = \bigcup_{(i,j)\in\mathcal{P}(\mathcal{S})} \; \bigcup_{(a,b)\in\mathcal{A}_{\text{intra}}\cup\mathcal{A}_{\text{inter}}} \Gamma_{ij}^{a,b} \tag{13}$$

where $\Gamma_{ij}^{a,b}$ denotes the latent segment connecting support states (x_i^a, y_i^a) and (x_j^b, y_j^b) under the local model $F_{ij}^{a\leftrightarrow b}$. Crucially, $\mathcal{M}_{\text{PFS}}$ is not a separate global network; it is the graph-induced union of local latent segments learned from $\mathcal{S}$. Operationally, ViPS-FSL uses $\mathcal{M}_{\text{PFS}}$ as the sampling domain for virtual cohort decoding:

$$\mathcal{V} = \mathcal{G}(\mathcal{S}) = \text{Decode}\big(\text{Sample}(\mathcal{M}_{\text{PFS}})\big) \tag{14}$$

where $\text{Sample}(\mathcal{M}_{\text{PFS}})$ denotes graph-edge selection plus latent interpolation along a selected segment $\Gamma_{ij}^{a,b}$. To sample from $\mathcal{M}_{\text{PFS}}$, we first select a graph-indexed latent segment $\Gamma_{ij}^{a,b} \subset \mathcal{M}_{\text{PFS}}$. The corresponding source and target states are encoded as $z_{\text{src}} = \text{Enc}_{\theta_{ij}}^{(a,b)}(x_{\text{src}}, y_{\text{src}})$ and $z_{\text{tgt}} = \text{Enc}_{\theta_{ij}}^{(a,b)}(x_{\text{tgt}}, y_{\text{tgt}})$. An initial latent trajectory is generated by spherical linear interpolation (SLERP) [22]:

$$z_k = \text{SLERP}(z_{\text{src}}, z_{\text{tgt}};\, \omega_k), \quad \omega_k \in [0, 1] \tag{15}$$

As SLERP weights do not ensure uniform anatomical variation, we calibrate interpolation by computing mask-derived volumes and fit a B-spline to map ω_k to volume, yielding calibrated weights ω_k^* with quasi-uniform volume change along the latent path. This constrains interpolated states to follow a physiologically plausible volume trajectory between the ED and ES anchors, rather than arbitrary appearance blending:

$$z_k^* = \text{SLERP}(z_{\text{src}}, z_{\text{tgt}};\, \omega_k^*) \tag{16}$$

The calibrated codes are decoded into virtual image–mask pairs; aggregating the local subsets over all edges and phase-pair types forms the full cohort $\mathcal{V} = \mathcal{G}(\mathcal{S})$ (Eq. 3). Because decoding is joint in image–mask space, every sample is labeled and couples cross-pathology variation with phase dynamics. The segmentation network (nnU-Net [2]) is trained on the labeled support augmented with $\mathcal{V}$, using no unlabeled data, pseudo-labeling, or query-time adaptation; the real support $\mathcal{S}$ also fits the co-translation models and defines the PFS. This addresses (B1) via graph-based densification, (B2) by relying solely on labeled support ($\mathcal{V} = \mathcal{G}(\mathcal{S})$, no unlabeled data), and (B3) by dense sampling of intermediate pathology–phase states absent from $\mathcal{S}$.

3 Experiments and Results

Datasets. We use ACDC [1] (100 subjects: healthy controls plus four pathology groups) and M&Ms [23] (320 subjects; multi-center, multi-vendor, multi-pathology). Support sets of 7 and 11 subjects are drawn from the 70-subject ACDC training split, stratified across pathologies ($\geq$1 per pathology, each contributing ED and ES). We evaluate in-distribution on a disjoint ACDC test split (N=50) and out-of-distribution on the full unseen M&Ms dataset (320 subjects). The single-source cross-domain protocol follows established single-source domain-generalization practice in cardiac MRI [24].

Strict Evaluation Protocol. For a strictly one-to-one comparison, ViPS-FSL and all five baselines are trained and evaluated on identical data. All methods use the same support-set and cross-validation patients following [12] under our same preprocessing (cropping, normalization), and all baselines are re-trained from scratch. All baselines additionally consume 63/59 unlabeled ACDC subjects, whereas ViPS-FSL uses none.

Implementation Details. We resampled all data to 1.5×1.5 mm and center-cropped to 256×256×16. Co-translation models were trained on all support-graph and phase pairs (1000 epochs, Adam, lr=1e−4); $\lambda_{\text{reg}}, \lambda_{\text{fid}} = 1$, $\lambda_{\text{mask}} = 20$, $\lambda_{\text{sm}} = 40$, $\lambda_{\text{LPIPS}} = 1$, $\lambda_{\text{SSIM}} = 10$, $\lambda_{\text{mSSIM}} = 1$, $\lambda_{\text{NCC}} = 10$ (Eq. 9–12). Volume-calibrated SLERP (200 steps/segment) yielded 4/6.4k labeled 3D samples for the 7/11 regimes, decoded into cohort $\mathcal{V}$; an nnU-Net [2] was trained on the augmented support set (200 epochs). Models are independent and trained in parallel (∼10 min on one H100); the 11-subject regime uses 220 models (∼37 GPU-h), ∼4 GPU-h for decoding and nnU-Net training.

Virtual-Cohort Distribution Fidelity. As the densified intermediate states have no paired ground truth, we validate fidelity at the distribution level: for each method (stochastic elastic augmentation, a vanilla GAN, and our cohort) we generated 10K 2D samples from the same 11 subjects and compared them to a held-out ACDC split (N=50) via KID [25]. Against the floor (real support, 0.023), our virtual patients reach 0.207: 4× closer (better) than elastic augmentation and 9× closer than the GAN (Table 1). ViPS-FSL yields faithful samples (Supplementary Fig. A); decoded samples (Fig. 2) interpolate smoothly between real endpoints.

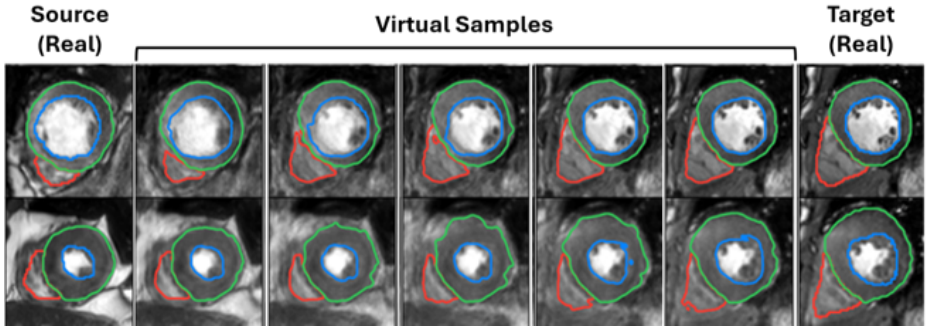

Fig. 2. Virtual patient samples. Top: cross-phase ED→ES ($\mathcal{A}_{\text{inter}}$); bottom: fixed-phase ($\mathcal{A}_{\text{intra}}$).

Table 1. Virtual cohort fidelity compared to unseen ACDC split (N=50).

Method	Reference Distribution	KID (↓)
Standard Augmentation (+Elastic Deform.)	Unseen ACDC (N=50)	0.887±0.008
GAN (trained on 11 subjects)		1.825±0.006
ViPS-FSL (Virtual Patients Cohort)		**0.207±0.006**
Support set (11 subjects) – Baseline		0.023±0.001

Table 2. In-distribution results on ACDC test split (N=50), for the 7/11-label regimes. Best per column in bold.

		7-label regime		11-label regime	
Method	Type	Dice (%) ↑	HD95 (mm) ↓	Dice (%) ↑	HD95 (mm) ↓
nnU-Net	Fully-sup. Baseline	45.6	23.6	49.9	20.6
SS-Net [13]	Consistency	79.8	8.5	83.3	6.6
MC-Net [14]	Self-train	80.0	8.4	83.9	6.4
BCP [12]	Consistency + Self-train	84.5	5.9	84.7	5.9
PPC [11]	Prototype	85.1	6.2	86.6	5.0
KnowSAM [17]	SAM distillation	85.9	6.8	87.6	4.2
ViPS-FSL (Ours)	Virtual-patients	**88.5**	**4.8**	**89.0**	**4.3**

In-Distribution Performance. On held-out unseen ACDC test (N=50), ViPS-FSL ranks first at both 7 and 11 support subjects with 88.5 and 89.0 average Dice, respectively, surpassing the strongest baseline by +2.6/+1.4 Dice, despite using no unlabeled data, while all baselines consume an additional 63/59 unlabeled subjects from the ACDC train split (Table 2).

Cross-Domain Generalization. We report strict single-source cross-domain performance on M&Ms in Table 3, comparing ViPS-FSL against five state-of-the-art FSL segmentation baselines (SS-Net [13], MC-Net [14], BCP [12], PPC [11], and KnowSAM [17]) spanning consistency regularization, self-training, and foundation-model distillation paradigms. All baselines use the remaining ACDC subjects as unlabeled streams (63/59); ViPS-FSL uses none, relying only on labeled support and the densified cohort. It attains the best overall Dice

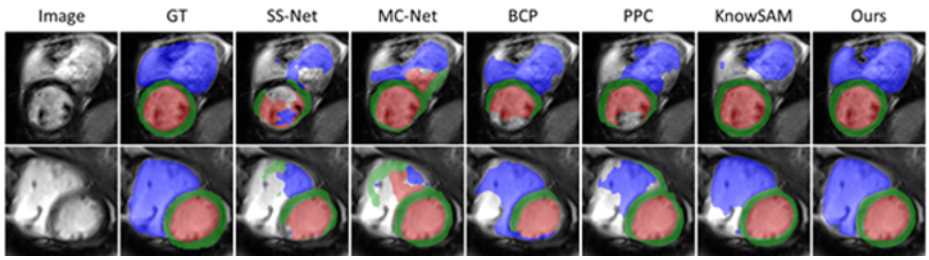

Fig. 3. Qualitative segmentation results. Our method vs. SOTA methods on M&Ms.

Table 3. Cross-domain comparison (M&Ms) using 7- and 11-label schemes. Method types: Consistency (Con.), Self-train (ST.), SAM distillation (SAM-KD), Virtual-patients (VP). Best per column in bold.

		#Subj. (ACDC)		Dice (%) ↑				HD95 (mm) ↓			
Method	Type	lab.	unlab.	RV	Myo	LV	Avg	RV	Myo	LV	Avg
nnU-Net	Full-Sup.	7	0	34.0	31.9	54.4	40.1	27.83	23.12	28.84	26.60
nnU-Net		11	0	46.5	25.9	53.6	42.0	27.06	23.21	25.84	25.37
SS-Net [13]	Con.		63	72.8	72.6	86.6	77.3	11.42	6.88	5.55	7.95
MC-Net [14]	ST.		63	73.1	77.0	88.8	79.6	11.45	5.10	4.61	7.05
BCP [12]	Con.+ST.	7	63	77.5	76.1	90.3	81.3	9.93	5.62	3.59	6.38
PPC [11]	Proto.		63	83.1	78.8	90.3	84.1	6.97	4.02	3.40	4.80
KnowSAM [17]	SAM-KD		63	82.2	82.7	90.6	85.2	8.90	3.11	3.90	5.30
ViPS-FSL (Ours)	VP		0	**87.7**	**83.7**	**90.9**	**87.4**	**5.34**	**2.50**	**2.93**	**3.59**
SS-Net [13]	Con.		59	74.2	76.1	86.7	79.0	11.40	5.27	6.16	7.61
MC-Net [14]	ST.		59	75.9	79.2	89.1	81.4	10.11	4.27	4.44	6.27
BCP [12]	Con.+ST.	11	59	79.2	74.8	89.0	81.0	9.89	6.02	4.30	6.74
PPC [11]	Proto.		59	83.1	80.8	91.1	85.0	7.69	3.60	3.05	4.78
KnowSAM [17]	SAM-KD		59	84.0	83.6	91.3	86.3	7.64	2.95	3.62	4.74
ViPS-FSL (Ours)	VP		0	**88.9**	**84.1**	**91.3**	**88.1**	**4.65**	**2.43**	**2.79**	**3.29**

and HD95 in both regimes: versus the best-performing baseline per-chamber, +5.5/1.0/0.3 Dice (RV/Myo/LV) with 23/20/14% lower HD95 at 7 labels, and +4.9/0.5 Dice (RV/Myo; LV matched) with 39/18/9% lower HD95 at 11 labels. Qualitative comparison (Fig. 3) further highlights ViPS-FSL's cleaner chamber boundaries than the baselines under shift.

Cross-Vendor Generalization. To assess per-vendor generalizability, we report single-source cross-domain results on M&Ms split by vendor—Siemens (A), Philips (B), GE (C), Canon (D). ViPS-FSL achieves the best Dice and lowest HD95 on every vendor: +4/2/2/1 Dice points and 32/28/20/14% lower HD95 at 7 labels, and +3/1/2/2 points with 34/28/16/17% lower HD95 at 11 labels, versus the strongest per-vendor baseline. Despite Siemens-only support (ACDC), it generalizes robustly to the unseen vendors (Philips/GE/Canon).

Ablation Study. We studied different phase-aware edge types (Table 5): intra-phase edges densify within-phase neighborhoods, while inter-phase edges span ED–ES variation. Combining both gives the best Dice and HD95 at 7 labels and

Table 4. Per-vendor cross-domain generalization on M&Ms. Best per column in bold.

		Dice (%) ↑				HD95 (mm) ↓			
Method	#labels	A	B	C	D	A	B	C	D
SS-Net [13]	7	74	80	81	73	8.9	6.9	6.2	10.4
MC-Net [14]		75	83	82	78	8.6	5.8	5.9	8.3
BCP [12]		80	83	84	76	6.6	5.3	4.7	10.3
PPC [11]		82	86	86	82	5.6	4.0	4.1	6.0
KnowSAM [17]		82	87	86	86	6.6	4.6	4.8	5.1
Ours		**86**	**89**	**88**	**87**	**3.8**	**2.9**	**3.3**	**4.4**
SS-Net [13]	11	76	81	82	73	8.9	6.9	6.2	10.4
MC-Net [14]		79	83	84	80	7.4	5.5	5.1	7.3
BCP [12]		79	83	84	76	7.3	5.7	4.5	10.5
PPC [11]		83	87	86	83	5.6	3.9	3.8	6.4
KnowSAM [17]		84	88	87	86	5.7	4.0	4.7	4.8
Ours		**87**	**89**	**89**	**88**	**3.7**	**2.8**	**3.2**	**4.0**

Table 5. Ablation for phase-aware edge typing schemes. Best per column in bold.

Intra	Inter	#labels	Dice (%) ↑	HD95 (mm) ↓
✓		7	87.08	3.870
	✓		87.20	3.911
✓	✓		**87.50**	**3.593**
✓		11	88.03	3.401
	✓		**88.15**	3.318
✓	✓		**88.15**	**3.295**

the best HD95 at 11 labels, confirming that the two edge types are complementary, especially in low-data regimes.

4 Conclusions

We presented ViPS-FSL, which reframes few-shot cardiac segmentation as support-set densification: instead of relying on unlabeled data, it decodes a fully labeled virtual-patient cohort from a graph-induced feature space to augment the support set. This tackles support-set underutilization, unlabeled-data dependency, and the cross-domain gap. Results show that ViPS-FSL ranks first for in-distribution segmentation and, transferring from single-source supervision to the unseen multi-vendor M&Ms cohort, outperforms five baselines that all rely on unlabeled data. In conclusion, densifying the labeled support is an effective route to domain-robust few-shot cardiac segmentation. We plan to apply the method to other datasets and modalities.

Acknowledgments. This work was supported in part by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH) under Grant R01HL169780.

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

References

1. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans on Medical Imaging. **37**(11), 2514–2525 (2018)
2. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods. **18**, 203–211 (2021)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T., and Nishi-no, K. (eds.) Computer Vision – ECCV 2022 Workshops. pp. 205–218. Springer Nature Swit-zerland, Cham (2023).
4. Cheng, Z., Wang, S., Xin, T., Zhou, T., Zhang, H., Shao, L.: Few-Shot Medical Image Seg-mentation via Generating Multiple Representative Descriptors. IEEE Trans on Medical Imag-ing. **43**, 2202–2214 (2024)
5. Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., Rueckert, D.: Self-supervised learning for few-shot medical image segmentation. IEEE Trans on Medical Imaging. **41**(7), 1837–1848 (2022)
6. Tang, S., Yan, S., Qi, X., Gao, J., Ye, M., Zhang, J., Zhu, X.: Few-shot medical image segmen-tation with high-fidelity prototypes. Medical Image Analysis. **100**, 103412 (2025)
7. Zhao, Z., Zhou, F., Zeng, Z., Guan, C., Zhou, S.K.: Meta-hallucinator: Towards Few-Shot Cross-Modality Cardiac Image Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., and Li, S. (eds.) MICCAI 2022, pp. 128–139. Springer Nature Switzerland, Cham (2022)
8. Fan, C., Jia, X., Xiao, A., Yu, H., Yang, Z., Yang, D., Xu, H., Huang, Y., Wang, L.: SPENet: Self-guided Prototype Enhancement Network for Few-Shot Medical Image Segmentation. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., and Park, J. (eds.) MICCAI 2025, pp. 584–593. Springer Nature Switzerland, Cham (2026)
9. Liu, Y., Xiao, H., Chai, J., Zhang, Y., Wang, R., Meng, Z., Luo, Z.: SynPo: Boosting Training-Free Few-Shot Medical Segmentation via High-Quality Negative Prompts. In: Gee, J.C., Alex-ander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., and Park, J. (eds.) MICCAI 2025, pp. 594–603. Springer Nature Switzerland, Cham (2026)
10. Kim, H., Hansen, S., Kampffmeyer, M.: Tied Prototype Model for Few-Shot Medical Image Segmentation. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkata-raman, A., Golland, P., Kim, J.H., and Park, J. (eds.) MICCAI 2025, pp. 651–661. Springer Nature Switzerland, Cham (2026)
11. Yuan, Y., Wang, X., Yang, X., Heng, P.-A.: Effective Semi-Supervised Medical Image Seg-mentation With Probabilistic Representations and Prototype Learning. IEEE Trans on Medical Imaging. (2024).
12. Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y.: Bidirectional copy-paste for semi-supervised medical image segmentation - IEEE/CVF, pp. 11514–11524 (2023)

13. Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring Smoothness and Class-Separation for Semi-supervised Medical Image Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., and Li, S. (eds.) MICCAI 2022, pp. 34–43. Springer Nature Switzerland, Cham (2022)
14. Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L.: Semi-supervised Left Atrium Segmentation with Mutual Consistency Training. In: De Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., and Essert, C. (eds.) MICCAI 2021, pp. 297–306. Springer International Publishing, Cham (2021)
15. Hu, J., Zhou, T., Huang, K., Zhou, Y., Zhang, H., Fan, B., Fu, H.: Uncertainty-guided Proto-type Reliability Enhancement Network for Few-Shot Medical Image Segmentation. IEEE Trans on Medical Imaging. (2025).
16. Wang, K., Zhan, B., Zu, C., Wu, X., Zhou, J., Zhou, L., Wang, Y.: Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learn-ing. Medical Image Analysis. **79**, 102447 (2022)
17. Huang, K., Zhou, T., Fu, H., Zhang, Y., Zhou, Y., Gong, C., Liang, D.: Learnable Prompting SAM-Induced Knowledge Distillation for Semi-Supervised Medical Image Segmentation. IEEE Transactions on Medical Imaging. **44**(6), 2295–2306 (2025)
18. Pham, C.M., Nguyen, P.L., Nguyen, T.T., Phan, V.M.H., Nguyen, B.P.: Unleashing SAM for Few-Shot Medical Image Segmentation with Dual-Encoder and Automated Prompting. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., and Park, J. (eds.) MICCAI 2025, pp. 670–680. Springer Nature Switzerland, Cham (2026)
19. Zhang, Y., Yang, J., Liu, Y., Cheng, Y., Qi, Y.: SemiSAM: Enhancing semi-supervised medical image segmentation via SAM-assisted consistency regularization. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 3982–3986. IEEE (2024)
20. Kong, L., Lian, C., Huang, D., Hu, Y., Zhou, Q.: Breaking the dilemma of medical image-to-image translation. Advances in Neural Information Processing Systems. **34**, 1964–1978 (2021)
21. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 586–595. IEEE, Salt Lake City (2018)
22. White, T.: Sampling Generative Networks, arXiv:1609.04468 (2016)
23. Campello, V.M., et al. Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge. IEEE Trans Med Imaging. **40**(12), 3543–3554 (2021)
24. Chen, C., Li, Z., Ouyang, C., Sinclair, M., Bai, W., Rueckert, D.: MaxStyle: Adversarial Style Composition for Robust Medical Image Segmentation. In: Medical Image Computing and Computer Assisted Intervention, In: MICCAI 2022. LNCS, vol. 13435, pp. 151–161. Springer, Cham (2022)
25. Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying MMD GANs. In: Int. Conf. on Learning Representations (ICLR) (2018)