Reviving Poor Object Segmentations in OOD Medical Images using Variational-Deep-PCA Modeling on Segmentation Maps with Sampling-Free Learning

Jimut B. Pal[1]*, Shantanu Welling[2], Himali Saini[2], Suyash P. Awate[1,2]
[1]Centre for Machine Intelligence and Data Science, Indian Institute of Technology (IIT) Bombay
[2]Computer Science and Engineering Department, Indian Institute of Technology (IIT) Bombay

Abstract

For object segmentation in medical images, deep neural networks (DNNs) typically perform poorly on out-of-distribution *(OOD) images stemming from the large variability in image-acquisition equipment and protocols across sites. However, we observe that the variability in the underlying object-*segmentation maps *is far lower. Thus, we propose a novel DNN framework to model this variability in segmentation maps, and leverage it to revive poor segmentations produced by existing DNNs on OOD images. Our DNN framework (i) learns the* principal modes of variation *in a class of segmentation maps, (ii) models each segmentation map using a low-dimensional mixture-of-modes latent representation on a simplex, (iii) enables* sampling-free variational learning *and* uncertainty *estimation, and (iv) trains using small in-distribution image sets. In special cases when OOD-image segmentations are extremely poor, we propose a* human-in-the-loop *method needing minuscule human intervention. Results using 6 publicly-available datasets and 8 existing DNN segmenters show the benefits of our framework in OOD-image object segmentation.*

1. Introduction

Deep neural networks (DNNs) for segmenting anatomical objects in medical images face severe challenges in *out-of-distribution* (OOD) images [9, 33] (unavailable during DNN training) that can arise from the large variability in image-acquisition equipment and protocols across clinical sites. Noting that the variability in object geometry is lower, we propose a novel DNN framework to model this variability in a class of *segmentation maps* by learning the *principal modes of variation*, and leverage it to revive poor segmentations produced by existing DNNs on OOD images; "segmentation map" is an image where pixel values are the probability of that pixel belonging to the object of interest.

Generative DNNs model *distributions* on latent variables, e.g., many probabilistic and Bayesian versions of principal component analysis (PCA) [6, 31], their nonlinear extensions [5], and variational autoencoders (VAEs) [18]. Early versions of PCA were restricted to multivariate-Gaussian based models, and nonlinear extensions required hand-crafting features or kernels. In contrast, VAEs leverage DNNs to learn a rich set of features from data, but require expensive Monte-Carlo sampling during learning and inference. A very recent kernel-based PCA method uses hierarchical/"deep" modeling [32], but avoids DNN models, avoids variational models, and focuses on industrial process fault detection (an application much different from ours). While some early works use a Bayesian interpretation of the softmax layer to model *uncertainty* in segmentation [15], we leverage it to propose (i) a novel *variational* DNN model enabling *sampling-free* learning and inference, and (ii) a novel efficient method to improve poor segmentations by optimizing plausible ones.

This paper focuses on improving poor segmentations of OOD images. A class of DNN methods [11, 14, 20] use *anatomical information* through loss terms designed on segmentations in the domain of interest, but they are inapplicable to our problem setting dealing with OOD data that is unavailable during training. Some other methods aim to improve poor segmentations using statistical models of shape (distance transforms or pointsets), but need expert annotations on OOD data during inference [13] or computationally intractable alignment between OOD images and shape models [28]; our framework avoids shape spaces. Recent methods leverage convexity-based models [10, 23, 24, 34] to improve segmentation of OOD images. Some recent methods [22] aim to simulate millions of anatomically-feasible segmentations and then find the one closest to a poor segmentation, but acknowledge that simulation can be unreliable without expert judgement and the inference-time search is expensive; our framework uses its learned model to efficiently optimize a feasible segmentation. Last, in cases of extremely poor OOD-image segmentations, if the object has star-convex [29] geometry, we propose to improve segmentations by obtaining *sparse segmentation*

*Supported by the Prime Minister's Research Fellowship.

maps with *minuscule human input* and then "projecting" them onto our learned principal modes.

2. Related Work

Existing DNNs for medical image segmentation include several variants of UNet [27] employing attention gates [16, 21, 25], residual connections [16, 37], and squeeze-and-excitation blocks and spatial pyramidal pooling [7, 16]. BASNet [26] uses hybrid-loss and residual refinement modules for saliency prediction and refinement, aiming to produce sharper and clearer boundaries. DSTransUNet [19] uses self-attention as in vision transformers [8], and hierarchical swin transformers in its U-shaped architecture to extract coarse and fine-grained features of different semantic classes. SegAN [36] uses an adversarial critic network with multiscale L1 losses, and captures long-range and short-range spatial relationships through local and global features. MedSegDiff [35] uses denoising diffusion probabilistic models [12] with dynamic conditional encoding for step-wise attention and feature-frequency parsing to reduce effects of high-frequency noise during diffusion.

A recent method [22] on statistical shape modeling uses a constrained VAE to learn a representation on valid cardiac shapes, then empirically sample millions of shapes and save only those (around 4 million) that pass an anatomical-correctness test, and replace poor segmentations by searching for their nearest anatomically correct saved shape. Unlike their strategy that requires large storage, expensive search, and a non-trivial domain-specific method of verifying anatomical correctness, our framework avoids any storage, quickly optimizes for the closest segmentation map using our DNN decoder, and enables sampling-free variational learning and uncertainty estimation. Some approaches train a DNN to restore (quality-enhance) poor segmentation maps but, unlike our framework, their DNN needs annotations of anatomical landmarks during inference [11] (unavailable for OOD data) and is unable to estimate uncertainty. A recent method [13] uses shape modeling in the target/alternate domain conditioned on a carefully selected set of 9 anatomical landmarks. During inference on cross-domain data, the landmarks are either (i) provided by an expert, but who may be unavailable or expensive, or (ii) detected by another DNN, but that requires training on OOD data that is unavailable by definition. Its very recent extension [14] replaces landmark information with edge information but also requires training on cross-domain data.

3. Proposed Method

Our VarDeepPCA framework learns a nonlinear statistical model of variability in segmentation maps for a class of objects in medical images. First, we propose VarDeepPCA's variational encoder-decoder model for a distribution of segmentation maps comprising (i) a decoder modeling the nonlinear principal modes of variation, and their mixtures, in the spatial domain, (ii) a low-dimensional latent variable modeling the mixture proportions underlying a given segmentation map, (iii) an encoder mapping a given segmentation map to its latent representation, and (iv) variational/distribution modeling using sampling-free learning. Second, we propose a way to use VarDeepPCA to improve poor segmentations, given by existing DNNs, for OOD images either automatically or using little human intervention.

Image X models an *acquired medical image* comprising an object of interest. Image Y models an associated *expert segmentation map* (binary or fuzzy); our framework can easily handle multiple expert segmentation maps Y for a single X. Image W models the associated *unknown true segmentation map*, which may differ from each associated Y. $\Phi(\cdot)$ models an *existing DNN segmenter* mapping images X to their object segmentations $\Phi(X)$. $\widetilde{X}$ models an OOD image for which the segmentation $\widetilde{Z} := \Phi(\widetilde{X})$ is of poor quality (Figure 1(A)). During training $\Phi(\cdot)$, OOD images and their expert segmentations are both unavailable.

3.1. DNN-based PCA Model on Segmentation Maps Using a Latent Representation on a Simplex

The VarDeepPCA framework can employ generic encoder-decoder DNN architectures to model a distribution on segmentation maps Y for a class of objects (Figure 1(B)). We assume the distribution comprises K *principal (nonlinear) modes of variation*, where K is small (this paper chooses $K := 8$), which are indicated by a K-length *one-hot* random vector C, where the index of the non-zero component indicates a specific mode of variation. Let $\mathbb{I}_k$ be a K-length one-hot vector with a value of 1 at index k. A typical segmentation map will be well-represented by a *mixture* of these K modes. We design VarDeepPCA to encode each segmentation Y using a *latent*-vector representation of K probabilities of Y being close to each of the K modes of variation, i.e., the vector $L := [P(C = \mathbb{I}_1|Y), \cdots, P(C = \mathbb{I}_K|Y)]$. Because L is a K-dimensional vector with positive values summing to 1, it lies on a $(K-1)$-dimensional *simplex* in $\mathbb{R}^K$. To get the latent representation L, VarDeepPCA's encoder $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$, parameterized by $\theta^{\mathcal{E}}$, first maps input Y to a K-dimensional *segmentation-feature* vector $S := \mathcal{E}(Y;\theta^{\mathcal{E}})$. Subsequently, we propose to map S to the vector L using the *softmax* function because this softmax mapping implicitly performs variational/distribution modeling (shown in Section 3.2) and enables sampling-free variational learning (shown in Section 3.3). VarDeepPCA's decoder $\mathcal{D}(\cdot;\theta^{\mathcal{D}})$, parameterized by $\theta^{\mathcal{D}}$, maps latent representations L to output segmentation maps $W := \mathcal{D}(L;\theta^{\mathcal{D}})$.

3.2. Variational Interpretation of Softmax Mapping

VarDeepPCA reinterprets the softmax function, underlying the mapping $L := \text{Softmax}(S)$, using Bayesian princi-

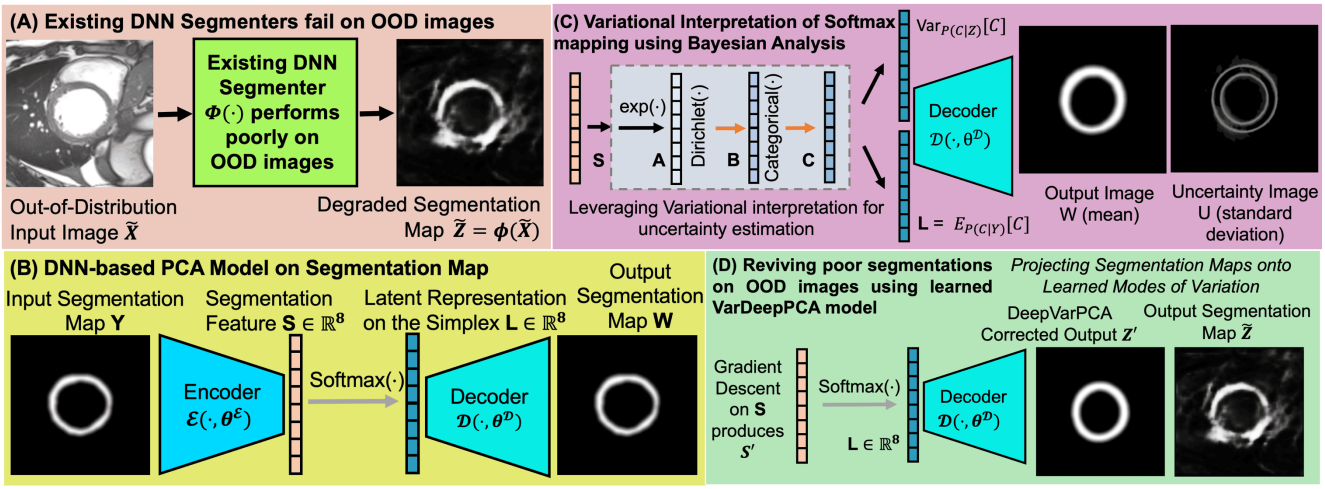


Figure 1. **VarDeepPCA on Segmentation Maps with Sampling-Free Variational Learning.** **(A)** Existing DNN-based object segmenters typically perform poorly on OOD images. **(B)** Our VarDeepPCA models and learns principal modes of variation of segmentation maps and a related latent representation on a simplex (Section 3.1) through a softmax mapping (Section 3.2). **(C)** Our Bayesian interpretation of the softmax endows a variational model with (i) closed-form marginalization enabling sampling-free variational learning (Section 3.3) and (ii) enabling per-pixel uncertainty estimates (Section 3.4). **(D)** Automatically reviving poor segmentations (Section 3.5(A)), produced by existing DNNs on OOD images, by "projecting" segmentation maps onto the VarDeepPCA's model of principal modes of variation.

ples to shed light on the underlying variational/distribution modeling on the simplex in $\mathbb{R}^K$ (Figure 1(C)). We model $P(C|S)$ as the *posterior-predictive distribution* arising from a *Categorial*-distribution *likelihood* $P(C|\cdot)$, on the modes of variation indicated by C, coupled with a *Dirichlet*-distribution (conjugate) *prior* $P(\cdot|S)$. Let the random vector A have its components $A_k := \exp(S_k) > 0$, for all $1 \leq k \leq K$, such that A parameterizes a Dirichlet distribution $\text{Dir}(B; A)$ of a hidden random vector B residing on the $(K-1)$-dimensional simplex. Because the mapping from Y to S to A is deterministic, the following equivalence between posterior-predictive distributions holds: $P(C|Y = y) \equiv P(C|S = \mathcal{E}(y; \theta^{\mathcal{E}})) \equiv P(C|A = \exp(\mathcal{E}(y; \theta^{\mathcal{E}})))$. Consider a categorical distribution $\text{Cat}(C; B)$ on one-hot vectors C, parameterized by the hidden random vector B that is sampled from its conjugate distribution $\text{Dir}(B; A)$. Then, the posterior-predictive distribution becomes

$$P(C|A) = \int_b P(C|b)P(b|A)db = \int_b \text{Cat}(C; b)\text{Dir}(b; A)db$$
$$= \int_b \left(\prod_k (b_k)^{C_k}\right)\left(\prod_k \frac{(b_k)^{A_k-1}}{\eta(A)}\right)db \quad (1)$$
$$= \int_b \prod_k \frac{(b_k)^{C_k+A_k-1}}{\eta(A)}db = \frac{\eta(A+C)}{\eta(A)}, \quad (2)$$

where the normalizing constant for the Dirichlet distribution is $\eta(F) := \prod_k \Gamma(F_k)/\Gamma(\sum_k F_k)$, where $\Gamma(\cdot)$ is the well-known Gamma function. This leads to

$$P(C|A) = \frac{\Gamma(\sum_k A_k)}{\prod_k(\Gamma(A_k))}\frac{\prod_k(\Gamma(A_k + C_k))}{\Gamma(\sum_k A_k + C_k)}. \quad (3)$$

Considering a specific instance of $c := \mathbb{I}_k$, and using the property $\Gamma(g+1) := g\Gamma(g)$ for gamma functions, simplifies the posterior-predictive distribution as

$$P(C = \mathbb{I}_k|A) = \frac{A_k}{\sum_{k=1}^K A_k} = \frac{\exp(S_k)}{\sum_{k=1}^K \exp(S_k)} = L_k \quad (4)$$

that is the k-th component of the vector $L = \text{Softmax}(S)$. Furthermore, the expected value $E_{P(C|Y)}[C]$ indeed equals L that we designed as the low-dimensional latent representation for Y. So, while the softmax mapping from S to the latent representation L is deterministic, the softmax implicitly (i) subsumes variational modeling by modeling distributions $P(B|\exp(S)) \equiv P(B|A)$ and $P(C|B)$, (ii) then marginalizes out the random variable B leveraging Bayesian inference to produce the *analytically exact* posterior-predictive distribution $P(C|S)$ in *closed form*, and (iii) finally takes the expectation of C under the posterior-predictive distribution to give $L = E_{P(C|S)}[C]$ in *closed form*; the closed-form expressions enable sampling-free variational learning as described next in Section 3.3.

3.3. Sampling-Free Variational Learning using Closed-Form Marginalization with Softmax

Let the training set of N segmentation maps be $\{Y_n\}_{n=1}^N$. For input Y, VarDeepPCA models a distribution $P(C|Y)$ representing a mixture of the modes of variation underlying Y. We use Bayesian decision theory to formulate the variational learning to maximize, over parameters θ, the *expected utility* (under $P(Y)$) of the decoder's output that is associated with its input as the expected value

of C under $P(C|Y)$. To measure the quality of any decoder output W, we choose the *utility function* to be the soft Dice-similarity-coefficient (sDSC) between W and the reference Y. So VarDeepPCA's variational-learning formulation is

$$\max_{\theta^{\mathcal{E}},\theta^{\mathcal{D}}} E_{P(Y)}[\text{sDSC}(Y, \mathcal{D}(L = E_{P(C|Y;\theta^{\mathcal{E}})}[C]; \theta^{\mathcal{D}}))] \quad (5)$$

$$\equiv \max_{\theta^{\mathcal{E}},\theta^{\mathcal{D}}} \sum_{n=1}^{N} \text{sDSC}(Y_n, \mathcal{D}(\text{Softmax}(\mathcal{E}(Y_n; \theta^{\mathcal{E}}); \theta^{\mathcal{D}}))) \quad (6)$$

where Section 3.2 derived the exact analytical value of (the expected value) L in closed-form using the softmax function. Thus, the VarDeepPCA learning formulation, despite involving a latent distribution $P(C|Y)$ (and distributions $P(C|B), P(B|Y)$) eliminates Monte-Carlo sampling, and the associated reparameterization, that becomes inevitable in typical variational DNNs (e.g., VAEs) because of the intractability of their underlying integration (Figure 1(C)).

3.4. Getting Output Segmentation with Uncertainty

For an input segmentation map Z, VarDeepPCA's internal low-dimensional representations (e.g., S and L in $\mathbb{R}^K$) are designed to model Z using only the top K modes of variation, while filtering out the remaining variation (e.g., arising from segmentation errors and discretization errors in Z). Also, for input Z, the variational model underlying VarDeepPCA outputs a latent distribution $P(C|Z) := P(C|B)P(B|Z)$ where $P(B|Z = z)$ is equivalent to $P(B|A = \exp(\mathcal{E}(z; \theta^{\mathcal{E}})))$. Indeed, VarDeepPCA can sample $c \sim P(C|Z)$ as follows: $S \leftarrow \mathcal{E}(Z; \theta^{\mathcal{E}})$, $A \leftarrow \exp(S)$, $b \sim \text{Dir}(B; A)$, $c \sim \text{Cat}(C; b)$. For a trained VarDeepPCA (as per Section 3.3), for input Z, we propose to (i) map the mean of $P(C|Z)$ through the decoder to infer the *output segmentation* W, i.e., $W := \mathcal{D}(E_{P(C|Z)}[C]; \theta^{\mathcal{D}})$ where $E_{P(C|Z;\theta^{\mathcal{E}})}[C] = \text{Softmax}(\mathcal{E}(Z; \theta^{\mathcal{E}})) = L$, and (ii) map the variance of $P(C|Z)$ through the decoder to get the *variance image* V in the spatial domain (described next), and then take per-pixel square-root values in V to get the associated *uncertainty image* U (Figure 1(C)). For the i-th pixel in V, we evaluate the variance V_i using (a) the variances $P(C_k|Z)$ of each component C_k; for categorial $P(C|Z)$, these have the closed-form $L_k(1 - L_k)$; and (b) the Jacobian of the decoder mapping $\mathcal{D}(\cdot; \theta^{\mathcal{D}})$ evaluated at $L := E_{P(C|Z)}[C]$:

$$V_i := \sum_{k=1}^{K} \left(\frac{\partial \mathcal{D}_i(L)}{\partial L_k}\bigg|_{L=\text{Softmax}(\mathcal{E}(Z))} \right)^2 \text{Var}_{P(C|Z)}[C_k],$$

where $\mathcal{D}_i(L)$ denotes the i-th pixel value in $\mathcal{D}(L)$.

3.5. Improving DNN Segmenters on OOD Images

We leverage the learned VarDeepPCA model in two different ways to revive poor segmentation maps $\widetilde{Z}$ produced by existing DNNs $\Phi(\cdot)$ on OOD images $\widetilde{X}$.

(A) Reviving Poor Segmentations Automatically. We propose a novel two-stage algorithm (Figure 1(D)). First, we pass $\widetilde{Z}$ through the encoder-decoder VarDeepPCA to filter out from $\widetilde{Z}$ the non-principal components of the segmentation map. This produces the filtered segmentation map $\overline{Z} := \mathcal{D}(\text{Softmax}(\mathcal{E}(\widetilde{Z}; \theta^{\mathcal{E}})); \theta^{\mathcal{D}})$. Second, we explicitly "project" $\overline{Z}$ onto the learned space of principal modes of variation by (i) fixing $\overline{Z}$ as the output reference, (ii) optimizing the segmentation-feature vector $S^* \in \mathbb{R}^K$ as $S^* := \arg\max_S \text{sDSC}(\overline{Z}, \mathcal{D}(\text{Softmax}(S); \theta^{\mathcal{D}}))$, using gradient descent, and (iii) defining the improved/revived segmentation $W^* := \mathcal{D}(\text{Softmax}(S^*); \theta^{\mathcal{D}})$.

(B) Reviving Very Poor Segmentations: Human in the Loop. Very poor segmentation maps $\widetilde{Z}$ cannot be revived well by the automatic strategy described earlier. In this case, we propose a novel two-stage algorithm (Figure 2) that requires very little human input and works well for objects defined by star-convex [29] boundaries, e.g., the myocardium is defined by its outer boundary (epicardium) and the inner boundary (endocardium) both of which are star-convex. First, making a reasonable assumption that the field of view in the OOD image $\widetilde{X}$ is similar to that in the training images X or Y, we ask the intervening human to roughly estimate the object centroid in $\widetilde{X}$, then center a polar coordinate system at that location, partition the polar domain into F sectors, and ask the human to indicate the correct segmentation of $\widetilde{X}$ in $f \ll F$ sectors by simply drawing two arcs corresponding to the outer and inner boundaries of the object; this segmentation $\mathring{Z}$ is incomplete and extremely sparse. Second, we explicitly "project" $\mathring{Z}$ onto the learned space of principal modes of variation in a way similar to that described earlier for automatic revival, but with one difference: to adapt to the sparsity of the reference segmentation map $\mathring{Z}$, we replace sDSC by *Partial-sDSC* that is computed using *only* the pixels lying within the f sectors that the human segmented, i.e.,

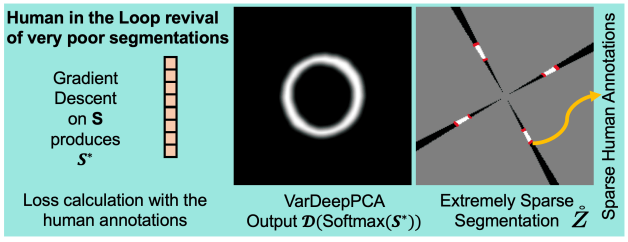


Figure 2. **Human-in-the-Loop Approach.** Reviving poor segmentations, produced by existing DNNs on OOD images, by using minuscule *human intervention* to obtain *extremely sparse segmentation maps* and "projecting" them onto the VarDeepPCA's model of principal modes of variation (Section 3.5(B)). For image on right: region in polar domain without any human annotation (gray); human-annotated arcs (red) in 4 (out of 72) sectors define the sparse segmentation (black ≡ background, white ≡ foreground); partial-sDSC is computed only in non-gray region.

Datasets → DNN Methods ↓	CAP			ACDC (OOD)			ACMRI (OOD)			MAGRABI			ORIGA (OOD)			G1020 (OOD)		
	Dice	HD95	ASD	Dice	HD95	ASD	Dice	HD95	ASD	Dice	HD95	ASD	Dice	HD95	ASD	Dice	HD95	ASD
UNet	90.1 (8.0)	5.4 (8.8)	1.5 (1.3)	60.6 (10.8)	41 (16)	11.6 (5.3)	71.3 (11.7)	26 (17)	6.9 (3.8)	85.9 (6.9)	9.9 (11.1)	4.1 (3.5)	82.1 (6.4)	16.0 (14.0)	5.5 (3.0)	78.9 (9.2)	15.6 (17)	6.2 (5.2)
UNet+VarDeepPCA (Ours)	89.2 (4.2)	3.7 (1.2)	1.5 (0.4)	79.7 (5.8)	5.2 (1.5)	2.1 (0.5)	79.8 (7.5)	6.0 (1.3)	2.5 (0.5)	91.1 (3.6)	5.6 (1.8)	2.2 (0.8)	89.6 (3.2)	6.5 (1.7)	2.3 (0.6)	88.9 (3.2)	6.5 (1.5)	2.6 (0.7)
AttnUNet	93.0 (3.8)	2.9 (3.8)	1.0 (0.5)	74.8 (7.8)	14 (10.5)	4.1 (2.0)	78.1 (9.0)	7.9 (5.4)	2.7 (1.0)	92.7 (3.9)	5.6 (5.5)	2.0 (1.4)	89.7 (3.9)	7.4 (7.1)	2.5 (1.5)	89.2 (5.0)	7.9 (9.2)	2.9 (2.6)
AttnUNet+VarDeepPCA (Ours)	90.2 (3.0)	3.4 (1.0)	1.4 (0.4)	81.5 (5.0)	5.2 (1.2)	2.2 (0.5)	82.3 (5.1)	5.6 (1.2)	2.3 (0.5)	93.6 (2.2)	4.4 (1.5)	1.6 (0.5)	90.4 (3.1)	5.9 (1.5)	2.1 (0.6)	91.4 (2.3)	5.5 (1.4)	2.0 (0.5)
ResUNet++	91.8 (5.8)	3.3 (2.9)	1.1 (0.7)	62.9 (14.7)	31 (17)	7.4 (4.3)	73.9 (11.8)	11.8 (9.2)	3.2 (1.7)	93.3 (2.4)	4.9 (3.4)	1.8 (0.9)	90.7 (3.0)	5.9 (1.4)	2.1 (0.5)	90.7 (3.3)	11.7 (18)	3.8 (4.3)
ResUNet+++VarDeepPCA (Ours)	89.6 (3.5)	3.6 (1.0)	1.4 (0.4)	76.4 (10.3)	5.8 (1.9)	2.2 (0.7)	79.3 (7.7)	6.1 (1.5)	2.5 (0.5)	93.9 (2.1)	4.2 (1.5)	1.5 (0.5)	90.9 (3.2)	5.8 (1.5)	2.0 (0.6)	92.1 (2.1)	5.3 (1.4)	1.9 (0.5)
DeepLabV3+	92.6 (3.2)	2.7 (1.3)	1.0 (0.4)	75.7 (8.8)	10.0 (5.5)	4.0 (2.2)	72.8 (10.8)	12 (10)	4.3 (2.9)	91.6 (3.1)	5.8 (2.1)	2.2 (0.8)	87.6 (4.3)	6.6 (1.7)	2.7 (0.8)	89.6 (3.1)	8.4 (7.6)	2.8 (1.5)
DeepLabV3++VarDeepPCA (Ours)	89.7 (3.2)	3.6 (1.0)	1.4 (0.4)	80.3 (6.0)	5.5 (1.1)	2.5 (0.7)	77.8 (6.4)	6.5 (1.3)	2.9 (0.7)	92.0 (2.9)	5.2 (1.8)	2.0 (0.7)	88.3 (4.3)	5.8 (1.6)	2.6 (0.8)	90.4 (2.7)	6.1 (1.4)	2.3 (0.6)
BASNet	79.9 (6.8)	6.8 (1.8)	3.0 (0.9)	74.6 (8.8)	7.5 (1.8)	3.4 (1.0)	75.8 (9.2)	7.6 (1.2)	3.6 (1.3)	93.9 (2.2)	4.4 (1.6)	1.6 (0.6)	91.4 (2.7)	5.3 (1.5)	1.9 (0.5)	92.5 (2.2)	5.3 (4.0)	1.9 (1.2)
BASNet+VarDeepPCA (Ours)	81.0 (7.2)	6.3 (1.7)	2.8 (0.9)	75.9 (8.6)	6.7 (1.6)	3.2 (1.0)	76.7 (8.7)	7.0 (1.7)	3.5 (1.2)	93.9 (2.1)	4.3 (1.6)	1.5 (0.5)	92.0 (3.0)	5.0 (1.6)	1.8 (0.5)	92.7 (2.2)	4.9 (1.5)	1.8 (0.5)
DSTransUNet	92.5 (3.8)	2.9 (1.7)	1.0 (0.6)	68.7 (9.8)	15.5 (6.5)	5.3 (1.9)	80.1 (6.2)	9.3 (6.7)	3.4 (1.6)	93.9 (3.0)	4.9 (7.0)	1.8 (2.1)	90.3 (3.1)	5.6 (4.3)	2.3 (1.2)	89.8 (4.8)	18.4 (22)	5.6 (6.2)
DSTransUNet+VarDeepPCA (Ours)	88.7 (3.4)	4.3 (1.4)	1.7 (0.4)	75.0 (7.2)	7.3 (1.4)	3.5 (0.9)	83.0 (4.8)	5.9 (1.2)	2.5 (0.6)	94.5 (1.8)	3.7 (1.3)	1.4 (0.4)	91.7 (3.1)	4.3 (1.2)	1.8 (0.6)	92.9 (2.0)	4.7 (1.5)	1.7 (0.5)
SegAN	93.7 (3.4)	2.3 (1.2)	0.9 (0.5)	67.4 (10.3)	38 (14)	11 (4.6)	77.0 (8.8)	11 (12)	4.0 (2.8)	94.3 (2.2)	5.2 (7.5)	1.9 (1.7)	89.7 (3.1)	9.6 (9.6)	3.3 (2.1)	91.5 (2.7)	13.3 (18)	4.4 (5.3)
SegAN+VarDeepPCA (Ours)	88.5 (3.4)	4.4 (1.4)	1.7 (0.4)	78.5 (6.2)	6.2 (1.4)	2.7 (0.7)	82.2 (5.3)	6.0 (1.3)	2.6 (0.7)	94.7 (2.0)	3.8 (1.5)	1.3 (0.4)	90.8 (3.2)	6.1 (1.7)	2.0 (0.6)	92.5 (2.3)	5.2 (1.7)	1.8 (0.5)
MedSegDiff	81.1 (4.0)	5.2 (2.0)	3.2 (0.6)	69.6 (7.0)	9.4 (2.7)	4.9 (1.2)	77.6 (8.0)	9.6 (5.0)	4.1 (1.5)	88.1 (9.1)	7.4 (3.4)	2.9 (1.1)	86.0 (7.3)	7.3 (2.3)	3.1 (1.1)	87.3 (5.7)	7.1 (2.3)	3.2 (1.2)
MedSegDiff+VarDeepPCA (Ours)	81.4 (4.8)	6.5 (1.4)	3.1 (0.6)	70.7 (7.1)	8.7 (1.2)	4.4 (0.8)	77.0 (7.0)	8.3 (1.8)	3.9 (1.0)	89.4 (4.0)	6.9 (2.1)	2.7 (1.0)	87.0 (5.3)	7.0 (2.2)	3.0 (1.1)	87.9 (4.5)	6.8 (2.2)	3.0 (1.1)
Human in the Loop using ~5.6% annotations (Ours)	87.1 (3.8)	4.4 (1.5)	1.8 (0.5)	84.3 (5.3)	4.4 (1.4)	1.7 (0.5)	80.6 (5.2)	6.1 (1.4)	2.5 (0.6)	96.4 (1.2)	2.5 (2.8)	1.1 (0.9)	96.7 (1.2)	2.0 (1.1)	1.0 (1.0)	95.1 (1.8)	5.8 (13)	2.0 (3.4)

Figure 3. **Results. Myocardium** datasets: CAP used for training (in-distribution data), and ACDC and A-CMRI as OOD data. **Retina** datasets: Magrabi used for training (in-distribution), and ORIGA and G1020 as OOD data. Numbers indicate means (standard deviations). Colors indicate improvements that are statistically significant (t-test p-value < 0.05) over other corresponding DNNs (ours or baseline).

$S^* := \arg\max_S$ Partial-sDSC$(\mathring{Z}, \mathcal{D}(\text{Softmax}(S); \theta^{\mathcal{D}}))$.

4. Results and Discussion

Training Details. We use 6 publicly-available datasets and 8 existing DNN segmenters. We pre-process the image data by cropping/padding and resampling to 256×256 pixels. VarDeepPCA's architecture uses an encoder $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$ having a sequence of convolution and max-pooling layers (number of channels reduces by 2× each time from 256 to 8; image size reduces by 2× each time from 256×256 to 1×1) until it produces a 1×1×8 feature vector $S \in \mathbb{R}^8$, and a corresponding decoder $\mathcal{D}(\cdot;\theta^{\mathcal{D}})$ using transpose-convolutions for upsampling. We use batch normalization after each convolution layer and Adam [17] optimization.

Baselines. We compare with 8 existing DNN segmenters (denoted $\Phi(\cdot)$ earlier) spanning several kinds of architectures and formulations/losses over the last decade, i.e., UNet [27], AttnUNet [21], ResUNet++ [16], DeepLabV3+ [7], BASNet [26] (uses boundary-aware loss), SegAN [36] (uses discriminative learning), DSTransUNet [19] (uses dual swin transformers), and MedSegDiff [35] (uses diffusion-process modeling). We train all methods on one dataset (in each domain: cardiac, retina) and then test them on two OOD datasets from that domain comprising images from different sources. For a fair analysis, we compare each baseline with its version augmented with our VarDeepPCA (Section 3.5); VarDeepPCA relies on the same training data as each baseline; number of parameters in our VarDeepPCA model is only 35% of that in the smallest baseline model (ResUNet++) and 1% of that in the largest baseline model (DSTransUNet). Performance measurement uses (i) DSC and (ii) the distribution of inter-boundary (predicted versus ground-truth) distances (as done for Hausdorff distance), in pixel units, in terms of the 95-th percentile (termed HD95) and the average (termed ASD).

Datasets. We use 3 short-axis cardiac MRI datasets: CAP [30], ACDC [4], A-CMRI [2]. We use CAP for training, validation (to tune free parameters), and in-distribution

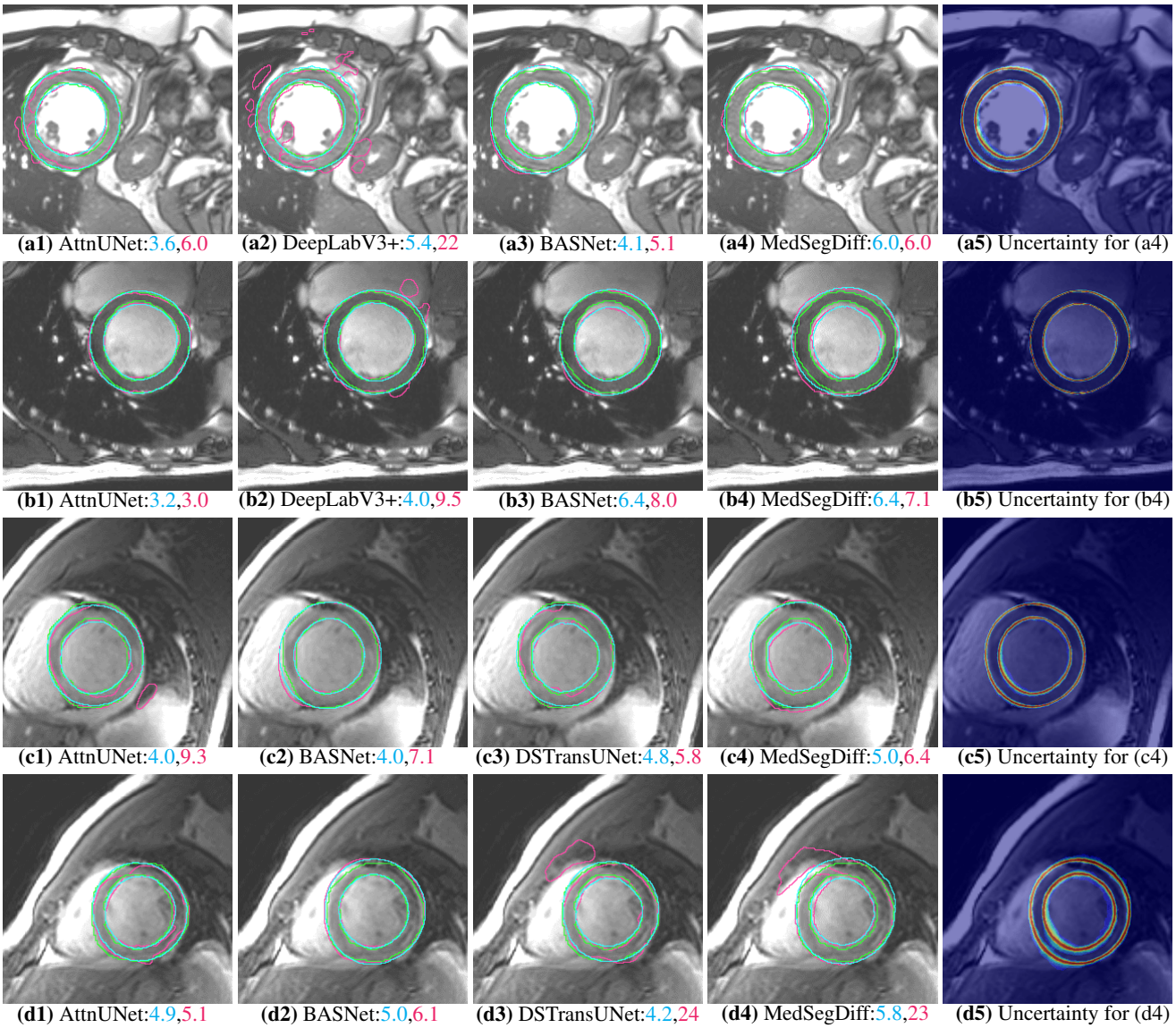


Figure 4. **Results (Myocardium): Automatic Revival on 4 best-performing baselines on OOD data. (a1)–(a4)** and **(b1)–(b4)** from ACDC dataset. **(c1)–(c4)** and **(d1)–(d4)** from ACMRI dataset. **(a5)–(d5)** Uncertainty maps produced using VarDeepPCA on top of the associated baselines in (a4)–(d4). Color scheme: Baseline; Baseline+VarDeepPCA (Ours); Ground Truth. Numbers indicate HD95 values.

testing; we use ACDC and A-CMRI for OOD testing. We use 3 retinal-image datasets: Magrabi [1], ORIGA [38], G1020 [3]. We use Magrabi for training, validation, and in-distribution testing; we use ORIGA and G1020 for OOD testing. For training all methods, to mimic a typical clinical scenario, we choose a small training set with 150 medical images along with their segmentation maps that are representative of the underlying distribution of segmentation maps. We use a separate validation set of 50 medical images (to tune hyperparameters of existing DNNs), and use the remaining images as in-distribution and OOD test sets.

Human-in-the-Loop Approach. We partition the polar domain into $F := 72$ sectors (each spanning 5 degrees) and get annotations of object boundaries (inner and outer) from a human intervener (Section 3.5) in $f := 4$ sectors randomly and uniformly spread across 360 degrees, needing only 5.6% of the effort needed for full object segmentation.

4.1. Results: Quantitative and Qualitative

Existing DNNs Segment OOD Images Poorly. Figure 3 demonstrates that all 8 baselines perform very well on in-distribution images, but quite poorly on OOD images; this exemplifies the major challenge faced by typical current DNN segmenters on applications across medical-imaging datasets. The poor segmentations on OOD images by existing DNNs often show gross errors, implausible object

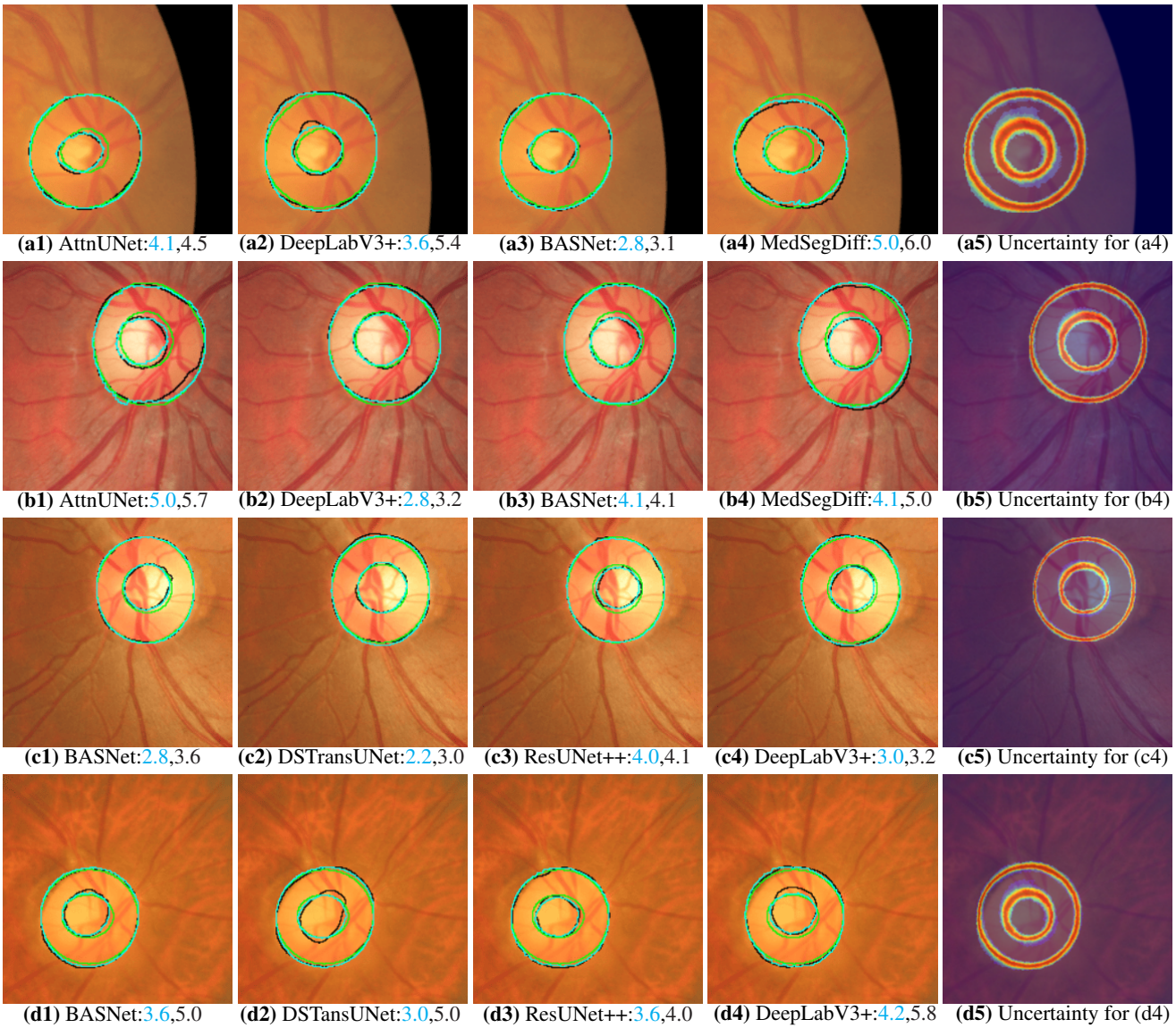
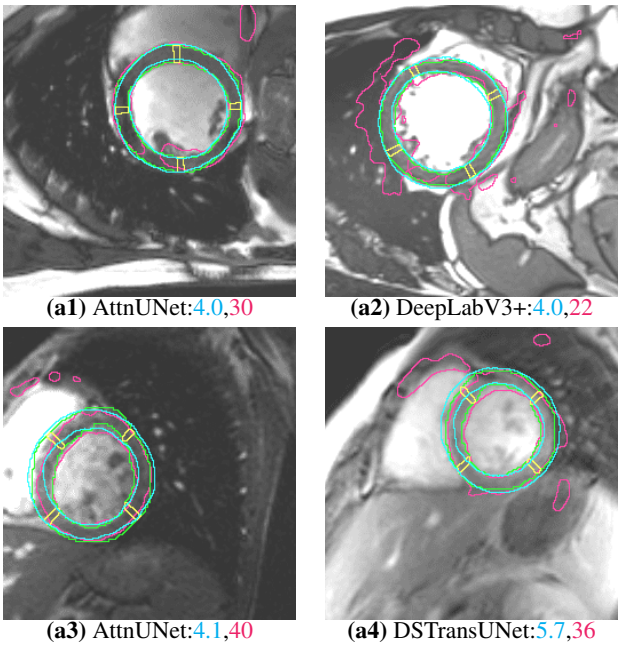


Figure 5. **Results (Retina): Automatic Revival on 4 best-performing baselines on OOD data. (a1)–(a4)** and **(b1)–(b4)** from G1020 dataset. **(c1)–(c4)** and **(d1)–(d4)** from ORIGA dataset. **(a5)–(d5)** Uncertainty maps produced using VarDeepPCA on top of the associated baselines in (a4)–(d4). Color scheme: Baseline; Baseline+VarDeepPCA (Ours); Ground Truth. The numbers indicate HD95 values.

shapes, and incorrect number of connected components in the object (Figure 4, Figure 5, Figure 6, Figure 7).

Automatic Revival of Poor Segmentations. On in-distribution images, the values of HD95 and ASD are small, in absolute terms, for all methods, and these results typically exhibit virtually imperceptible differences from the ground truth. On in-distribution images, VarDeepPCA-augmented versions (e.g., UNet+VarDeepPCA) perform at par with the underlying baselines (e.g., UNet): a bit worse for myocardium and a bit better for the retina (Figure 3). On OOD images, VarDeepPCA-augmented versions consistently outperform the underlying baselines statistically significantly, leading to huge improvements in all performance measures for, both, the myocardium and the retina, quantitatively (Figure 3) and qualitatively (Figure 4, Figure 5). Moreover, almost every VarDeepPCA-augmented version outperforms almost all the baselines on OOD images (Figure 3), e.g., ResUNet+++VarDeepPCA (with only 6M parameters) performs almost-always significantly better than DSTransUNet (171M parameters). In this way, the principal variability in segmentation maps learned by VarDeepPCA is able to filter-out errors in segmentations and improve poor segmentation maps by "projecting" them onto the principles modes of variation (Figure 4, Figure 5).

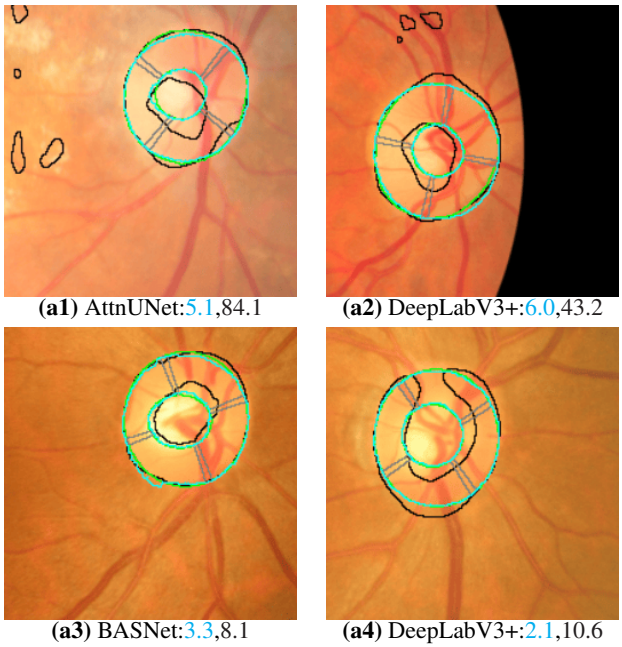
Uncertainty Images Associated with VarDeepPCA Segmentations. Unlike baselines, the variational modeling



(a1) AttnUNet:4.0,30 **(a2)** DeepLabV3+:4.0,22

(a3) AttnUNet:4.1,40 **(a4)** DSTransUNet:5.7,36

Figure 6. **Results (Myocardium): Human-in-the-Loop Revival on 2 best-performing baselines on OOD data** from **(a1)-(a2)** ACDC and **(a3)-(a4)** ACMRI datasets. Color scheme: Baseline; Human-in-the-Loop Annotations; Baseline+VarDeepPCA (Ours); Ground Truth. The numbers indicate HD95 values.



(a1) AttnUNet:5.1,84.1 **(a2)** DeepLabV3+:6.0,43.2

(a3) BASNet:3.3,8.1 **(a4)** DeepLabV3+:2.1,10.6

Figure 7. **Results (Retina): Human-in-the-Loop Revival on 2 best performing baselines on OOD data** from **(a1)-(a2)** G1020 and **(a3)-(a4)** ORIGA datasets. Color scheme: Baseline; Human-in-the-Loop Annotations; Baseline+VarDeepPCA (Ours); Ground Truth. The numbers indicate HD95 values.

underlying VarDeepPCA enables per-pixel uncertainty estimates. The uncertainty is high in object regions where the statistical model of segmentation-map variations indicates a larger variability. For example, compared to the optic-disc boundary (Figure 5(a5)–(d5)), higher uncertainty around the optic-cup boundary stems from a lower reliability in human/expert annotaters there because of lower contrast in the retinal image around the optic-cup boundary. Similarly, compared to the uncertainty in the retinal images, the uncertainty is lower around the myocardium (Figure 4(a5)–(d5)) because of (i) possibly lower anatomically variability and (ii) better contrast in cardiac MRI around the myocardium increasing the reliability of human/expert segmentations.

Human-in-the-Loop Revival of Very Poor Segmentations. Even with only 5.6% of the object segmented by the human annotater, the statistical model of segmentation-map variability learned by VarDeepPCA completes the extremely-sparse human segmentation very accurately (Figure 6, Figure 7). Indeed, VarDeepPCA is unaffected by the OOD variations in medical images (stemming from variations in image-acquisition equipment and protocols across sites) because it solely relies on segmentation maps for learning and analysis. Figure 3 clearly shows that our human-in-the-loop approach performs (i) always better than all other baselines, (ii) and almost-always better than the VarDeepPCA-augmented version.

5. Conclusion

We propose to revive poor segmentations produced by existing DNNs on OOD images using the novel VarDeepPCA framework that learns the principal modes of variation in segmentation maps; VarDeepPCA trains on a small in-distribution dataset, without using any OOD data. For each segmentation map Y, VarDeepPCA designs its latent representation $L = E_{P(C|Y)}[C]$ to model the mixture probabilities of each mode of variation being associated with the segmentation map. VarDeepPCA produces L using a softmax mapping, and shows that this softmax mapping implicitly performs variational modeling, enables computationally efficient sampling-free variational learning and inference. VarDeepPCA has a lightweight architecture (ResUnet+++VarDeepPCA with 6M parameters outperforms DSTransUNet with 171M parameters), is modality independent, leverages generic encoder-decoder archiectures, and produce uncertainty estimates. In cases of extremely poor OOD-image segmentations, if the object exhibits star-convex geometry, VarDeepPCA revives the segmentations by using minuscule human intervention to obtain extremely sparse segmentation maps and "projecting" them onto the VarDeepPCA's model of principal modes of variation. Results using 6 publicly-available datasets and 8 existing DNN segmenters show that VarDeepPCA outperforms existing DNNs in OOD-image object segmentation.

References

[1] A Almazroa, S Alodhayb, E Osman, E Ramadan, M Hummadi, M Dlaim, M Alkatee, K Raahemifar, and V Lakshminarayanan. Retinal fundus images for glaucoma analysis: the RIGA dataset. In *Medical Imaging*, volume 10579, pages 1–9, 2018. 6

[2] A Andreopoulos and J Tsotsos. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Medical Image Analysis*, 12(3):335–57, 2008. 5

[3] M Bajwa, G Singh, W Neumeier, M Malik, A Dengel, and S Ahmed. G1020: A benchmark retinal fundus image dataset for computer-aided glaucoma detection. In *Int. Joint Conf. on Neural Networks*, volume 1, pages 1–7, 2020. 6

[4] O Bernard, A Lalande, C Zotti, F Cervenansky, X Yang, P Heng, I Cetin, K Lekadir, O Camara, B Gonzalez, A Miguel, G Sanroma, S Napel, S Petersen, G Tziritas, E Grinias, M Khened, V Kollerathu, G Krishnamurthi, M Rohé, X Pennec, M Sermesant, F Isensee, P Jäger, K Maier-Hein, P Full, I Wolf, S Engelhardt, C Baumgartner, L Koch, J Wolterink, I Išgum, Y Jang, Y Hong, J Patravali, S Jain, O Humbert, and P Jodoin. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Trans. on Med. Imag.*, 37(11):2514–25, 2018. 5

[5] S Bernhard, S Alexander, and M Klaus-Robert. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319, 1998. 1

[6] C Bishop. Variational principal components. In *Int. Conf. on Artificial Neural Networks*, volume 1, pages 509–14, 1999. 1

[7] L Chen, Y Zhu, G Papandreou, F Schroff, and H Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Eur. Conf. Comput. Vis.*, volume 11211, pages 833–50, 2018. 2, 5

[8] A Dosovitskiy, L Beyer, A Kolesnikov, D Weissenborn, X Zhai, T Unterthiner, M Dehghani, M Minderer, G Heigold, S Gelly, J Uszkoreit, and N Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *Int. Conf. on Learning Representations*, pages 1–21, 2021. 2

[9] S Farquhar and Y Gal. What 'Out-of-distribution' is and is not. In *Adv. Neural Inform. Process. Syst. ML Safety Workshop*, pages 1–7, 2022. 1

[10] A Gaikwad, H Varma, and S Awate. Deep variational segmentation of topology-constrained object sets, with correlated uncertainty models, for robustness to degradations. In *IEEE Int. Conf. Image Process.*, pages 2195–99, 2023. 1

[11] J Gao, Q Lao, P Liu, H Yi, Q Kang, Z Jiang, X Wu, K Li, Y Chen, and L Zhang. Anatomically guided cross-domain repair and screening for ultrasound fetal biometry. *IEEE Journal of Biomedical and Health Informatics*, 27(10):4914–25, 2023. 1, 2

[12] J Ho, A Jain, and P Abbeel. Denoising diffusion probabilistic models. In *Adv. Neural Inform. Process. Syst.*, volume 574, pages 6840–51, 2020. 2

[13] A Jacob, P Sharma, and D Rueckert. Deep conditional shape models for 3D cardiac image segmentation. In *Statistical Atlases and Computational Models of the Heart. Regular and CMRxRecon Challenge Papers*, volume 14507 of *Lecture Notes in Computer Science*, pages 44–54. Springer, 2023. 1, 2

[14] A Jacob, P Sharma, and D Rueckert. DCSM 2.0: Deep conditional shape models for data efficient segmentation. In *IEEE Int. Symposium on Biomedical Imaging*, pages 1–4, 2024. 1, 2

[15] Rohit Jena and Suyash P Awate. A bayesian neural net to segment images with uncertainty estimates and good calibration. In *Int. Conf. on Inf. Process. in Medical Imaging*, pages 3–15, 2019. 1

[16] D Jha, P Smedsrud, M Riegler, D Johansen, T Lange, P Halvorsen, and H Johansen. ResUNet++: An advanced architecture for medical image segmentation. *2019 IEEE Int. Symposium on Multimedia*, pages 225–30, 2019. 2, 5

[17] D Kingma and J Ba. Adam: a method for stochastic optimization. In *Int. Conf. Learn. Represent.*, pages 1–15, 2015. 5

[18] D Kingma and M Welling. Auto-Encoding Variational Bayes. In *Int. Conf. Learn. Represent.*, pages 1–14, 2014. 1

[19] A Lin, B Chen, J Xu, Z Zhang, G Lu, and D Zhang. DS-TransUNet: Dual swin transformer U-Net for medical image segmentation. *IEEE Trans. on Instrumentation and Measurement*, 71:1–15, 2021. 2, 5

[20] O Oktay, E Ferrante, K Kamnitsas, M Heinrich, W Bai, J Caballero, S Cook, A Marvao, T Dawes, D O'Regan, B Kainz, B Glocker, and D Rueckert. Anatomically Constrained Neural Networks (ACNNs): Application to cardiac image enhancement and segmentation. *IEEE Trans. on Medical Imaging*, 37:384–95, 2017. 1

[21] O Oktay, J Schlemper, L Folgoc, M Lee, M Heinrich, K Misawa, K Mori, S McDonagh, N Hammerla, B Kainz, B Glocker, and D Rueckert. Attention U-Net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*, pages 1–10, 2018. 2, 5

[22] N Painchaud, Y Skandarani, T Judge, O Bernard, A Lalande, and P Jodoin. Cardiac segmentation with strong anatomical guarantees. *IEEE Trans. on Medical Imaging*, 39(11):3703–13, 2020. 1, 2

[23] J Pal and S Awate. Convex segments for convex objects using DNN boundary tracing and graduated optimization. In *Int. Conf. on Medical Image Computing and Computer Assisted Intervention*, pages 91–101, 2024. 1

[24] J Pal and S Awate. A hard convex-shape constraint in DNNs for object segmentation. In *IEEE Int. Conf. Image Process.*, pages 2074–80, 2024. 1

[25] J Pal and D Mj. Improving multi-scale attention networks: Bayesian optimization for segmenting medical images. *The Imaging Science Journal*, 71:33–49, 2023. 2

[26] X Qin, Z Zhang, C Huang, C Gao, M Dehghan, and M Jägersand. BASNet: Boundary-aware salient object detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7479–89, 2019. 2, 5

[27] O Ronneberger, P Fischer, and T Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Int.*

Conf. on Medical Image Computing and Computer-Assisted Intervention, volume 9351, pages 234–41. Springer, 2015. 2, 5

[28] S Shigwan, A Gaikwad, and S Awate. Object segmentation with deep neural nets coupled with a shape prior, when learning from a training set of limited quality and small size. In *IEEE Int. Symposium on Biomedical Imaging*, pages 1149–53, 2020. 1

[29] C Smith. A characterization of star-shaped sets. *American Mathematical Monthly*, 75:386, 1968. 1, 4

[30] A Suinesiaputra, B Cowan, A Al-Agamy, M Elattar, N Ayache, A Fahmy, A Khalifa, P Medrano-Gracia, M Jolly, A Kadish, D Lee, J Margeta, S Warfield, and A Young. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac mr images. *Medical Image Analysis*, 18(1):50–62, 2014. 5

[31] M Tipping and C Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 1999. 1

[32] F Tonin, Q Tao, P Patrinos, and J Suykens. Deep kernel principal component analysis for multi-level feature learning. *Neural Networks*, 170:578–95, 2024. 1

[33] D Tran, J Snoek, and B Lakshminarayanan. Practical uncertainty estimation and out-of-distribution robustness in deep learning. *Adv. Neural Inform. Process. Syst. Tutorial*, 2020. 1

[34] H Varma, A Gaikwad, and S Awate. Adversarial training with multiscale boundary-prediction DNN for robust topologically-constrained segmentation in ood images. In *IEEE Int. Symposium on Biomedical Imaging*, pages 1–5, 2023. 1

[35] J Wu, R Fu, H Fang, Y Zhang, Y Yang, H Xiong, H Liu, and Y Xu. MedSegDiff: Medical image segmentation with diffusion probabilistic model. In *Medical Imaging with Deep Learning*, volume 227 of *Proceedings of Machine Learning Research*, pages 1623–39, 2023. 2, 5

[36] Y Xue, T Xu, H Zhang, L Long, and X Huang. SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation. *Neuroinformatics*, 16:383–92, 2018. 2, 5

[37] Z Zhang, Q Liu, and Y Wang. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–53, 2018. 2

[38] Z Zhang, F Yin, J Liu, W Kee Wong, N Meng Tan, B Hai Lee, J Cheng, and T Yin Wong. ORIGA-light: An online retinal fundus image database for glaucoma analysis and research. *IEEE Int. Conf. Engineering in Medicine and Biology*, pages 3065–68, 2010. 6