Convex Segments for Convex Objects using DNN Boundary Tracing and Graduated Optimization

Jimut B. Pal and Suyash P. Awate

Centre for Machine Intelligence and Data Science, Indian Institute of Technology (IIT) Bombay, Mumbai, India

Abstract. Image segmentation often involves objects of interest that are biologically known to be *convex* shaped. While typical deep-neural-networks (DNNs) for object segmentation ignore object properties relating to shape, the DNNs that employ shape information fail to enforce *hard* constraints on shape. We design a brand-new DNN framework that *guarantees convexity* of the output object-segment by leveraging fundamental geometrical insights into the boundaries of convex-shaped objects. Moreover, we design our framework to build on typical existing DNNs for per-pixel segmentation, while maintaining simplicity in loss-term formulation and maintaining frugality in model size and training time. Results using six publicly available datasets demonstrates that our DNN framework, with little overheads, provides significant benefits in the *robust* segmentation of convex objects in *out-of-distribution* images.

Keywords: Object segmentation, deep neural network, convex segment, hard constraint, out-of-distribution, robustness, generalizability.

1 Introduction and Related Work

Biomedical image segmentation often involves an object/segment of interest that is *convex* shaped, e.g., (i) the regions bounded by the epicardium or by the endocardium in short-axis cardiac-MRI slices [7, 18, 23, 31] and (ii) the optic disc in retinal fundus images [29, 15]. Deep neural networks (DNNs) have been very effective in semantic image segmentation, e.g., several variants of UNet [20] employing attention gates [12, 17], residual connections [12, 34], and squeeze-and-excitation blocks and spatial pyramidal pooling [12]. Recently, [14] proposed a large DNN model adding hierarchical swin transformers within the encoder and the decoder of its UNet. Some DNNs [19, 10] use object-boundary related loss terms to improve predictions near object boundaries. However, none of these methods enforce any prior or constraint on object geometry or shape.

Some early segmentation methods impose hard shape constraints on segments [9, 21], but use hand-crafted features within graph-cut frameworks, unlike DNNs that learn complex and task-optimized image features. Some DNNs infuse shape priors within object segmentation, e.g., [16] promotes star-like shapes for skin-lesion segmentation, [33, 31] promote shape properties in the optic-disc

and epicardium segmentation, and [24] use a statistical shape prior for robust segmentation.

While these DNN-based shape priors can improve robustness to degradations, they fail to enforce any *hard constraint* on object geometry.

Some recent DNNs [11, 13, 25, 6] promote topological properties in segmentations by penalizing topological errors during training. TopoNet [11] works only on small patch sizes (65×65 pixels used by [11]), because its loss term relies on matching based on persistent homology, which, in their words, is "too complex and prone to errors". TeTrIS [13] starts with a candidate shape that is topologically correct and then deforms it to match the data, but this can be ineffective when the candidate shape has incorrect topology or when the deformation lead to incorrect topology. clDice [25] uses a loss term penalizing topological errors in tubular vessel structures. [6] train a DNN using a topological prior in terms of the Betti numbers, where the prior helps DNN learning in challenging tasks. *None* of these DNNs reliably enforce any hard topological or geometrical constraint. Some very recent work [30, 8] enforces star-like shapes but doesnot guarantee convexity.

Most of the methods for semantic segmentation evaluate their performance on a test set that has the same distribution as the training set. However, in biomedical imaging applications, at the time of deployment, the distribution of test images can differ from the distribution of training images because of several reasons. Such *out-of-distribution* (OOD) images [26, 36, 32] can occur because of variability in imaging equipment, clinical protocols, acquisition errors and artifacts, pathologies, and human errors. For such OOD images, typical DNNs lack *robustness* [22] to produce high-quality segmentations. Some methods like [24] incorporate shape information to improve segmentation quality for the applications where the training as well as the test images are noisy. However, such methods cannot guarantee object shape to have a specific geometry/shape property, e.g., convexity. In contrast, our framework leads to significant robustness to OOD variations while guaranteeing object-segment convexity.

This paper makes many contributions. We propose a novel DNN framework for segmenting a convex-shaped object in a 2D image, where our DNN parameterization guarantees the output object-segment to be convex. Our DNN design (i) leverages key geometrical insights into the boundaries of convex-shaped objects, (ii) can use image-based features from any existing DNN for per-pixel image segmentation, and (iii) maintains simplicity in loss-term formulation and maintains frugality in DNN size and training time. Results using six publicly available datasets show that our framework, with little overheads, provides significant benefits in the robust segmentation of convex objects in OOD images.

2 Proposed Method

Our DNN for object segmentation guarantees the convexity of its output segment using a novel parameterization. It leverages the geometric principle that the boundary of a *convex set*/segment is a *convex curve* [5]. We denote spatial coordinates in 2D by $Z := (X, Y)$ with abscissa $X \in \mathbb{R}$ and ordinate $Y \in \mathbb{R}$.

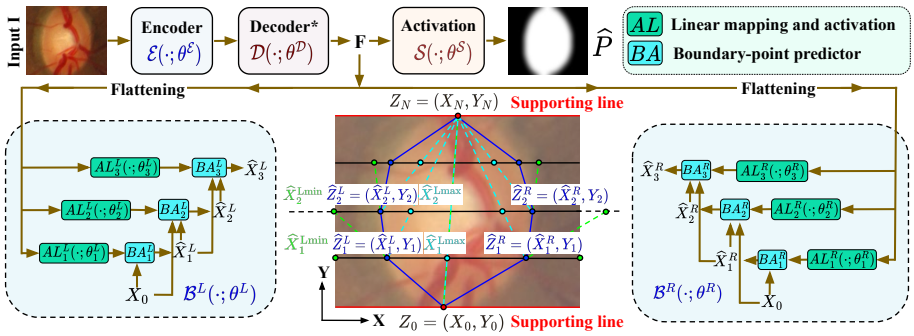

Fig. 1. Our DNN Design for Hard Convexity Constraint on Object Segment. ***Decoder**** is without the final activation layer.

2.1 Geometric Modeling of a Convex-Segment Boundary

The DNN models the entire object boundary (Figure 1) as a combination of two boundary parts, and then enforces segment-convexity related constrains on each part. These boundary parts are determined by any pair of parallel *supporting lines* [5] in the image domain, e.g., two horizontal supporting lines indicate the bottom-most and top-most points (Z_0, Z_N) on the object boundary, partitioning the boundary into a left and a right part. Such supporting lines can be easily provided by a user quickly providing the object's extremities in the vertical direction, or by a separate simple DNN trained for this task. If a supporting line touches the boundary over an interval where the boundary has zero curvature, then our DNN only models the boundary parts excluding this interval.

The DNN models the abscissas of the boundary parts on the left and right (X^L and X^R) as, respectively, convex and concave functions of the ordinate Y. The DNN models each boundary part as piecewise linear; in this paper, $N := 16$ pieces gave sufficient accuracy. Let the points on the left boundary, in sequence going from bottom to top, be $Z_0, Z_1^L, Z_2^L, \cdots, Z_N$. If we consider the directions of any ordered pair of line-segment vectors, e.g., $\overrightarrow{Z_0 Z_1^L}$ and $\overrightarrow{Z_1^L Z_2^L}$, then the direction of the second vector typically undergoes a clockwise rotation of the first, and never an anti-clockwise rotation. Similarly, while going from Z_0 to Z_N along the right-boundary, the curve typically turns anti-clockwise and never clockwise. Thus, starting with Z_0, along each boundary, our DNN *constrains each turn* so that (i) it is never in the wrong direction and (ii) the amount of turn is upper bounded to ensure the feasibility of all future turns. Given Z_0 and Z_N, we design two DNN components to produce estimates of the piecewise-linear boundaries on the left and right through $\{\widehat{Z}_i^L\}_{i=1}^{N-1}$ and $\{\widehat{Z}_i^R\}_{i=1}^{N-1}$, respectively.

2.2 DNN Architecture Guaranteeing Segment Convexity

We design a novel DNN architecture to estimate the boundary points on the left-boundary curve and the right-boundary curve, as modeled in Section 2.1, to guarantee the convexity of the resulting segment. Given the initial and final coordinates Z_0 and Z_N, we propose to estimate the coordinates for each curve in the

sequence $\widehat{Z}_1^L, \cdots, \widehat{Z}_{N-1}^L$ and $\widehat{Z}_1^R, \cdots, \widehat{Z}_{N-1}^R$ where the i-th coordinate's estimate on a curve ($\widehat{Z}_i^L$ or $\widehat{Z}_i^R$) is informed by the previously estimated coordinates (for $i' < i$) on that curve. Let the ordinates separating the N pieces on each curve be $Y_0 < Y_1 < \cdots < Y_{N-1} < Y_N$, which are equi-spaced and fixed. Then, we fix the ordinate estimates as $\widehat{Y}_i^L := \widehat{Y}_i^R := Y_i, \forall i$, and train the DNN to estimate the abscissas $\{\widehat{X}_i^L\}_{i=1}^{N-1}$ and $\{\widehat{X}_i^R\}_{i=1}^{N-1}$ *sequentially to guarantee segment convexity.*

For ordinate Y_0, let $\widehat{X}_0^L := \widehat{X}_0^R := X_0$. On the line segment $\overline{Z_0 Z_N}$, let the point with ordinate Y_1 have abscissa $\widehat{X}_1^{\text{Lmax}}$. Then, for ordinate Y_1, our DNN guarantees that its abscissa estimate $\widehat{X}_1^L$ is upper-bounded by $\widehat{X}_1^{\text{Lmax}}$ and lower-bounded by $\widehat{X}_1^{\text{Lmin}} := 0$ (where abscissa 0 corresponds to the image boundary on the left). For ordinate Y_2, given Z_0 and $\widehat{Z}_1^L$, the DNN guarantees that its estimate $\widehat{X}_2^L$ lies within the interval $(\widehat{X}_2^{\text{Lmin}}, \widehat{X}_2^{\text{Lmax}})$ where (i) $\widehat{X}_2^{\text{Lmin}}$ is the abscissa at the intersection of the extension of $\overrightarrow{Z_0 \widehat{Z}_1^L}$ with the horizontal line through ordinate Y_2 and (ii) $\widehat{X}_2^{\text{Lmax}}$ is the abscissa at the intersection of the line segment $\overline{\widehat{Z}_1^L Z_N}$ with the horizontal line through ordinate Y_2. Thus, we sequentially define $\widehat{X}_i^{\text{Lmin}} := f(\widehat{X}_{i-1}^L, \widehat{X}_{i-2}^L; Y_i, Y_{i-1}, Y_{i-2})$ for $2 \leq i \leq N-1$, and $\widehat{X}_i^{\text{Lmax}} := g(\widehat{X}_{i-1}^L; Z_N, Y_i, Y_{i-1})$ for $1 \leq i \leq N-1$, where functions $f(\cdot)$ and $g(\cdot)$ compute the aforementioned intersections as linear functions of their arguments.

To estimate the boundary points, the DNN leverages image-based features F (detailed next) and transforms them through a linear layer followed by an activation function, together modeled by $\text{AL}_i^L(F; \theta_i^L)$, with the output range $(0, 1)$ and parameterized by θ_i^L. Then, we scale and shift $\text{AL}_i^L(F; \theta_i^L)$ that lies within $(0, 1)$ to estimate the abscissa $\widehat{X}_i^L$ guaranteed to be within the interval $(\widehat{X}_2^{\text{Lmin}}, \widehat{X}_2^{\text{Lmax}})$. Let the function $\text{TS}(u; s, t) := su + t$ model scaling the input $u \in \mathbb{R}$ by $s \in \mathbb{R}$ and then applying a shift by $t \in \mathbb{R}$. Then, the DNN estimate of the abscissa is $\widehat{X}_i^L := \text{TS}(\text{AL}_i^L(F; \theta_i^L); \widehat{X}_i^{\text{Lmax}} - \widehat{X}_i^{\text{Lmin}}, \widehat{X}_i^{\text{Lmin}})$. Figure 1 combines the actions of the functions $f(\cdot)$, $g(\cdot)$, and $\text{TS}(\cdot)$ into a function that predicts the i-th boundary-point abscissa denoted by $\text{BA}_i^L(\text{AL}_i^L(F; \theta_i^L), \widehat{X}_{i-1}^L, \widehat{X}_{i-2}^L)$.

Analogously, for the right boundary, we define $\widehat{X}_0^R := X_0$, and sequentially estimate each boundary-point abscissa by first estimating its lower and upper bound as $\widehat{X}_i^{\text{Rmin}} := g(\widehat{Z}_{i-1}^R; Z_N, Y_i)$ where $1 \leq i \leq N-1$ and $\widehat{X}_i^{\text{Rmax}} := f(\widehat{Z}_{i-1}^R, \widehat{Z}_{i-2}^R; Y_i)$ where $2 \leq i \leq N-1$, and then estimating the abscissa itself as $\text{BA}_i^R(\text{AL}_i^R(F; \theta_i^R), \widehat{Z}_{i-1}^R, \widehat{Z}_{i-2}^R)$. Let $\theta^L := \cup_{i=1}^{N-1} \theta_i^L$ and $\theta^R := \cup_{i=1}^{N-1} \theta_i^R$.

We propose to design the image features F by leveraging the decoder of any existing probabilistic per-pixel segmenter DNN, e.g., a UNet. For the per-pixel segmenter, (i) let $\mathcal{E}(\cdot; \theta^{\mathcal{E}})$ model the encoder, parameterized by $\theta^{\mathcal{E}}$, mapping the input image to a latent vector; (ii) let $\mathcal{D}(\cdot; \theta^{\mathcal{D}})$, parameterized by $\theta^{\mathcal{D}}$, model a mapping from the latent vector to the features produced at the decoder's last layer; and (iii) let mapping $\mathcal{S}(\cdot; \theta^{\mathcal{S}})$, parameterized by $\theta^{\mathcal{S}}$, model an activation function producing the final per-pixel probabilities. For the boundary-estimator DNN component, let $\mathcal{B}(\cdot; \theta^{\mathcal{B}})$, parameterized by $\theta^{\mathcal{B}} := \theta^L \cup \theta^R$, map the decoder features F to the left and right boundary abscissas. Let $\theta := \theta^{\mathcal{E}} \cup \theta^{\mathcal{D}} \cup \theta^{\mathcal{S}} \cup \theta^{\mathcal{B}}$. For input image I, (i) the features are $F := \mathcal{D}(\mathcal{E}(I; \theta^{\mathcal{E}}); \theta^{\mathcal{D}})$; (ii) the segmenter's

output probability-map is $\widehat{P}(I;\theta) := \mathcal{S}(F;\theta^{\mathcal{S}})$; and (iii) the convex-boundary output by the boundary predictor has abscissa estimates $\widehat{X}(I;\theta) := \mathcal{B}(F;\theta^{\mathcal{B}})$.

2.3 Graduated Optimization Formulation and Strategy

The training set comprises J triples $\mathcal{T} := \{(I^j, P^j, X^j)\}_{j=1}^{J}$, where I^j is an input image, P^j is its per-pixel (binary) segmentation image, and X^j is the vector of abscissas of all boundary points (left and right both) for the chosen grid-ordinate values. The training minimizes the sum of the loss functions for the probabilistic per-pixel segmentation and the boundary estimation as $\mathcal{L}(\mathcal{T};\theta) := \sum_{j=1}^{J} \|\widehat{X}(I^j;\theta) - X^j\|_1 + \text{SoftDice}(\widehat{P}(I^j;\theta), P^j)$, where $\|\cdot\|_1$ is the L1 norm and SoftDice is the soft version of the Dice similarity coefficient. Training uses a 3-phase strategy: (i) train only the per-pixel segmenter using only the SoftDice loss to optimize $\theta^{\mathcal{E}} \cup \theta^{\mathcal{D}} \cup \theta^{\mathcal{S}}$; (ii) keeping $\theta^{\mathcal{E}} \cup \theta^{\mathcal{D}} \cup \theta^{\mathcal{S}}$ fixed, train only the boundary predictor using only the L1 loss to optimize $\theta^{\mathcal{B}}$; (iii) train the entire DNN over θ using both loss terms. While training the boundary predictor, we find that the following two-pass (graduated-optimization) training strategy leads to faster training and better results: (i) first relax the convexity constraint by scaling and shifting $\text{AL}_i^L(F;\theta_i^L) \in (0,1)$ to estimate the abscissa $\widehat{X}_i^L$ within a relaxed interval $(\widehat{X}_2^{\text{Lmin}} - \Delta, \widehat{X}_2^{\text{Lmax}} + \Delta)$, where $\Delta > 0$ is a small value compared to the interval length $|\widehat{X}_2^{\text{Lmax}} - \widehat{X}_2^{\text{Lmin}}|$ (we use Δ as 10% of the interval length), and (ii) then warm-start and enforce the convexity guarantee by setting $\Delta \to 0$.

3 Results and Discussion

Datasets. We use 3 short-axis cardiac MRI datasets: CAP [27], ACDC [4], A-CMRI [2]. We use the CAP dataset for training, validation (to tune free parameters), and testing (60%:15%:25% split); we use the ACDC and A-CMRI datasets for OOD testing, where in 50% of the images, we introduce slight Rician noise and slight spike artifacts [36]. We use 3 retinal-image datasets: Magrabi [1], ORIGA [35], G1020 [3]. We use the Magrabi dataset for training, validation, and testing (60%:15%:25% split); we use ORIGA and G1020 for OOD testing.

Methods. We compare with 9 existing methods: UNet [20], AttnUNet [17], ResUNet++ [34] (each using soft-Dice loss); UNet+DoU, AttnUNet+DoU, ResUNet+++DoU adding a boundary-based loss [28] to the UNet-based versions; UNet+Topo [6] adding a topology-promoting loss to the UNet; BASNet [19] with boundary-aware losses; DS-TransUNet [14] adding dual swin transformers to the UNet. We extend each of the first 3 methods using our framework guaranteeing segment convexity, to give cUNet, cAttnUNet, and cResUNet++; our methods take the final output from the convexity-guaranteeing boundary predictor (and not the segmenter). During the testing phase, we determine whether each pixel lies inside or outside the boundary of the boundary-predictor. Using the interior segment, we compute the Hausdorff-95 (HD95) and Average Surface Distance (HDavg) from the per-pixel segmentation. All methods train using Adam on images that are cropped (for a fair comparison) to a region limited by the left

Models ↓	CAP		**ACDC**	**A-CMRI**	Magrabi		**ORIGA**	**G1020**
	Time	mean (s.d.)	mean (s.d.)	mean (s.d.)	Time	mean (s.d.)	mean (s.d.)	mean (s.d.)
UNet	0.3	3.0 (1.4)	20 (4.9)	22 (5.8)	0.1	6.5 (2.5)	15.3 (2.8)	17.8 (4.2)
# params 7.8M	8.2	1.0 (0.4)	8.3 (2.2)	9.4 (2.8)	6.6	2.6 (1.1)	8.0 (1.6)	9.0 (2.3)
UNet+DoU	0.3	2.7 (2.1)	86 (22)	90 (22.8)	0.2	1.0 (0.3)	14.7 (4.8)	18 (6.8)
# params 7.8M	8.1	1.0 (0.6)	47 (11.7)	35 (11.9)	6.7	0.3 (0.2)	6.5 (1.6)	7.8 (2.7)
UNet+Topo	16.1	3.0 (1.4)	15.6 (3.2)	17.4 (4.3)	15.1	6.5 (2.5)	13.7 (1.1)	14.3 (2.7)
# params 7.8M	8.3	1.0 (0.4)	6.3 (2.1)	7.4 (1.9)	6.8	2.6 (1.1)	6.1 (0.7)	6.4 (1.2)
cUNet (Ours)	0.9	4.4 (1.8)	6.4 (1.6)	6.1 (1.4)	0.7	1.5 (0.4)	6.6 (2.3)	8.9 (2.76)
# params 9.2M	21	1.5 (0.6)	2.1 (0.5)	2.1 (0.5)	21.9	0.5 (0.0)	2.6 (0.8)	3.3 (0.9)
AttnUNet	0.4	3.8 (1.3)	9.0 (1.7)	9.6 (2.1)	0.3	7.3 (3.0)	13.7 (2.9)	17.2 (5.0)
# params 35M	15	1.3 (0.5)	3.4 (0.8)	3.6 (0.9)	11.4	2.9 (1.3)	7.1 (1.7)	8.6 (2.5)
AttnUNet+DoU	0.4	2.3 (1.1)	10.2 (3.0)	9.4 (2.4)	0.3	1.0 (0.0)	11.0 (2.2)	12.6 (3.8)
# params 35M	14	0.8 (0.3)	4.5 (1.3)	4.1 (1.0)	11.4	0.2 (0.0)	5.0 (1.2)	5.9 (1.7)
cAttnUNet (Ours)	1.7	4.0 (1.5)	6.5 (1.7)	6.4 (1.6)	1.9	2.3 (0.7)	5.5 (1.0)	7.4 (1.2)
# params 36M	33	1.3 (0.4)	2.2 (0.6)	2.2 (0.5)	27.9	0.9 (0.2)	2.2 (0.5)	3.0 (0.6)
ResUNet++	0.2	2.8 (1.0)	32 (8.0)	28 (7.5)	0.1	1.2 (0.4)	8.7 (2.6)	11.5 (4.3)
# params 4.1M	12	1.0 (0.4)	12.3 (3.0)	10.6 (2.8)	11.7	0.5 (0.0)	3.5 (1.1)	4.7 (1.7)
ResUNet+++DoU	0.3	2.2 (0.8)	40 (9.0)	37 (11)	0.2	1.0 (0.0)	6.3 (2.5)	11.7 (7.7)
# params 4.1M	13	0.8 (0.4)	17 (4.4)	13.9 (4.1)	12.5	0.3 (0.0)	2.5 (0.9)	4.6 (2.6)
cResUNet++ (Ours)	1.0	4.4 (1.6)	5.7 (1.0)	5.6 (1.0)	1.0	2.4 (0.7)	6.1 (2.7)	8.8 (3.0)
# params 5.5M	24	1.5 (0.6)	2.0 (0.4)	2.0 (0.4)	25.6	0.9 (0.2)	2.5 (1.1)	3.3 (1.0)
BASNet*	1.3	2.5 (1.1)	6.9 (1.5)	8.0 (2.2)	1.2	3.0 (1.6)	6.9 (1.6)	10.5 (2.3)
# params 174M	31	0.9 (0.4)	2.5 (0.6)	3.2 (0.9)	25.6	1.3 (0.8)	3.4 (1.0)	4.9 (1.3)
DS-TransUNet*	10	2.4 (1.1)	7.1 (1.7)	8.2 (2.2)	8.6	1.0 (0.0)	11.0 (2.2)	13.6 (3.0)
# params 171M	143	0.8 (0.3)	2.6 (0.6)	3.1 (0.9)	151	0.1 (0.0)	5.4 (1.4)	6.7 (1.8)

Fig. 2. Results: Epicardium in Cardiac MRI; Optic Disc in Retinal Images. In boxes in Time column, first row shows training time (hours); second row shows inference time per image (milliseconds). In other boxes, first row shows mean (and standard deviation) of HD95; second row shows those for HDavg; colored values (in blue) indicate a statistically significant improvement over all baseline methods based on a t-test. For BASNet and DS-TransUNet, * indicates pre-training with ImageNet1K.

and right extremities of the object (indicated by vertical parallel supporting lines provided to all methods). Performance measurement uses the distribution of inter-boundary (predicted-vs-true) distances (as done for Hausdorff distance), in pixel units, in terms of the average (termed HDavg) and the 95-th percentile (termed HD95).

Results: OOD Data. First, for OOD test images in cardiac MRI (ACDC, A-CMRI) and retinal images (ORIGA, G1020), our 3 methods (cUNet, cAttnUNet, cResUNet++), each guaranteeing a convex segment shape, greatly outperform their corresponding baselines (Figure 2; Figure 3, Figure 4), (i) in their original form (UNet, AttnUNet, ResUNet++) as well as (ii) their versions incorporating a boundary-based loss (UNet+DoU, AttnUNet+DoU, ResUNet+++DoU), with only one exception indicated in light blue (Figure 2) where ResUNet+++DoU performs close to our cResUNet++. Second, the 6 baseline versions of UNet, AttnUNet, ResUNet++, UNet+DoU, AttnUNet+DoU, and ResUNet+++DoU all typically lead to implausible object shapes with large errors, unlike our convexity-guaranteeing DNNs (cUNet, cAttnUNet, cResUNet++). Third, tuning UNet+Topo's free parameter (weighting likelihood and prior) using in-distribution images, where UNet itself gives good topologically-correct

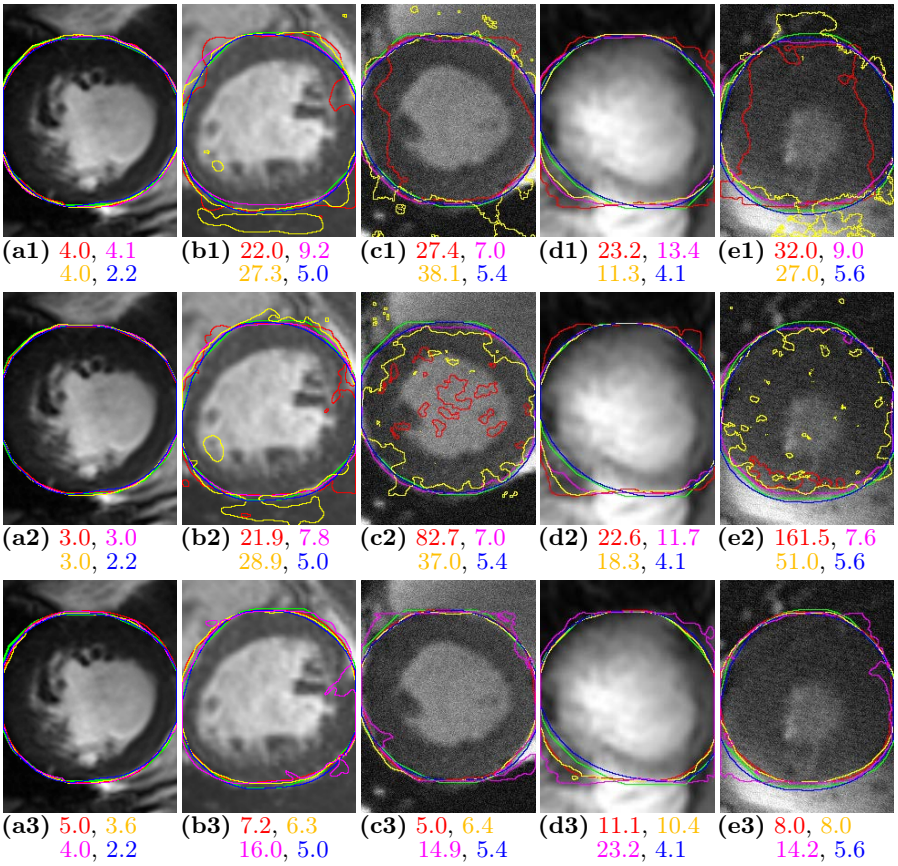

Fig. 3. Results: Epicardium. **(a1)**–**(a3)** In-distribution test image from CAP dataset. **(b1)**–**(b3)** OOD test image from **ACDC** dataset. **(c1)**–**(c3)** OOD test image after slight-degradation on **ACDC** dataset. **(d1)**–**(d3)** OOD test image from **A-CMRI** dataset. **(e1)**–**(e3)** OOD test image after slight-degradation on **A-CMRI** dataset. All images show color-coded segment boundaries from cResUNet++ (ours) and ground-truth. **(a1)**–**(e1)** also show segment boundaries from UNet, AttnUNet, ResUNet++. **(a2)**–**(e2)** also show segment boundaries from UNet+DoU, AttnUNet+DoU, ResUNet+++DoU. **(a3)**–**(e3)** also show segment boundaries from BASNet, UNet+Topo, DS-TransUNet. HD95 values are shown below, colored consistently.

results, is ineffective for OOD data. So, in this paper, UNet+Topo results for OOD data use parameter tuning on an OOD validation set, which is unfair to all other methods; even so, while UNet+Topo improves over UNet on OOD data, it takes 18× longer to train, still leads to implausible shapes, and is unable to improve over cUNet. Fourth, *each one* of our 3 methods (cUNet, cAttnUNet, cResUNet++) almost always greatly outperforms *every one* of the aforementioned 7 UNet-based baselines. Fifth, our 3 methods also improve over BASNet and DS-TransUNet despite the latter two using (i) 5-32× larger DNNs, (ii) pre-

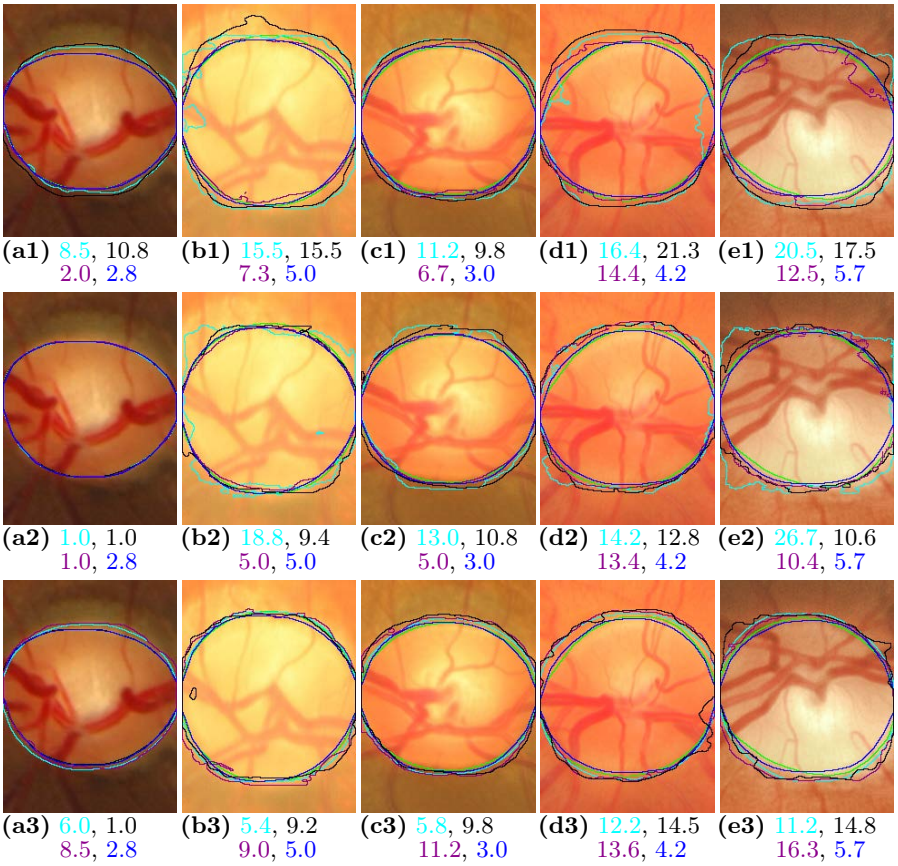

Fig. 4. Results: Optic Disc. (a1)–(a3) In-distribution test image from Magrabi dataset. **(b1)–(b3)** and **(c1)–(c3)** Two OOD test images from **ORIGA** dataset. **(d1)–(d3)** and **(e1)–(e3)** Two OOD test images from **G1020** dataset. All images show color-coded segment boundaries from cResUNet++ (ours) and ground-truth. **(a1)–(e1)** also show segment boundaries from UNet, AttnUNet, ResUNet++. **(a2)–(e3)** also show segment boundaries from UNet+DoU, AttnUNet+DoU, ResUNet+++DoU. **(a3)–(e3)** also show segment boundaries from BASNet, DS-TransUNet, UNet+Topo. HD95 values are shown below, colored consistently.

training using the large ImageNet1K dataset, and (iii) 6-10× more time to train; DS-TransUNet is also 4-6× slower during testing. As expected, BASNet and DS-TransUNet do improve over the 7 UNet-based baselines and mostly (but not always) lead to plausible and close-to-convex shapes.

Results: In-distribution Data. For CAP and Magrabi test images, all 9 existing methods perform comparably and statistically better than our 3 methods, but our errors remain small in magnitude while guaranteeing convexity.

Conclusion. This paper proposes a DNN framework for segmenting a convex-shaped object in a 2D image. It leverages fundamental properties of the bound-

aries of convex segments, and designs a DNN regressor that is parameterized to output a boundary corresponding to a convex segment. Our framework can leverage image-based features from any existing DNN for per-pixel image segmentation, maintains simplicity in loss-term formulation, and maintains frugality in DNN size. A minor limitation of our work is that the boundary-prediction block increases training and testing times compared to baseline models. Additionally, creating supporting lines at test time, which requires another DNN, introduces further complexity. To ensure a fair comparison, this method is used across all approaches. However, our per-image test time remains within 30 milliseconds on an Nvidia RTX2080Ti GPU, enabling real-time processing. Results using 6 publicly available datasets and 9 baseline methods show that our DNN framework provides significant benefits in the robust segmentation of convex objects in OOD images.

Acknowledgments. Supported by the Center for Machine Intelligence and Data Science (CMInDS) fellowship and the Prime Minister's Research Fellowship.

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

References

1. Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., Hummadi, M., Dlaim, M., Alkatee, M., Raahemifar, K., Lakshminarayanan, V.: Retinal fundus images for glaucoma analysis: the riga dataset. In: Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications. vol. 10579, p. 105790B (2018)
2. Andreopoulos, A., Tsotsos, J.: Efficient and generalizable statistical models of shape and appearance for analysis of cardiac mri. Medical Image Analysis **12**(3), 335–357 (2008)
3. Bajwa, M.N., Singh, G.A.P., Neumeier, W., Malik, M.I., Dengel, A., Ahmed, S.: G1020: A benchmark retinal fundus image dataset for computer-aided glaucoma detection. In: IJCNN. pp. 1–7 (2020)
4. Bernard, O., Lalande, A., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Trans. Med. Imag. **37**(11), 2514–25 (2018)
5. Busemann, H.: The geometry of geodesics. Elsevier Science (1955)
6. Clough, J., Byrne, N., Oksuz, I., Zimmer, V., Schnabel, J., King, A.: A topological loss function for deep-learning based image segmentation using persistent homology. IEEE Trans. Pattern Anal. Mach. Intell. **44**(12), 8766–78 (2022)
7. Epstein, F.: MRI of left ventricular function. J. Nucl. Cardiol. **14**(5), 729–44 (2007)
8. Gaikwad, A.V., Varma, H., Awate, S.P.: Deep variational segmentation of topology-constrained object sets, with correlated uncertainty models, for robustness to degradations. In: 2023 IEEE International Conference on Image Processing (ICIP). pp. 2195–2199 (2023)
9. Gulshan, V., Rother, C., Criminisi, A., Blake, A., Zisserman, A.: Geodesic star convexity for interactive image segmentation. In: IEEE Comp. Vis. Pattern Recog. pp. 3129–36 (2010)

10. Hatamizadeh, A., Terzopoulos, D., Myronenko, A.: End-to-end boundary aware networks for medical image segmentation. In: Machine Learn. Med. Imag. pp. 187–194 (2019)
11. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. In: Neural Info. Proc. Sys. (2019)
12. Jha, D., Smedsrud, P., Riegler, M., Johansen, D., Lange, T., Halvorsen, P., Johansen, H.: ResUNet++: An advanced architecture for medical image segmentation. In: Proc. IEEE Int. Symp. Multimedia. pp. 225–30 (2019)
13. Lee, M., Petersen, K., Pawlowski, N., Glocker, B., Schaap, M.: TeTrIS: Template transformer networks for image segmentation with shape priors. IEEE Trans. Med. Imag. **38**(11), 2596–606 (2019)
14. Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G., Zhang, D.: DS-TransUNet: Dual swin transformer U-Net for medical image segmentation. IEEE Trans. Instr. Meas. **71**, 1–15 (2022)
15. Lu, S.: Accurate and efficient optic disc detection and segmentation by a circular transformation. IEEE TMI **30**, 2126–33 (2011)
16. Mirikharaji, Z., Hamarneh, G.: Star shape prior in fully convolutional networks for skin lesion segmentation. In: Med. Imag. Comput. Comp.-Assist. Interv. pp. 737–45 (2018)
17. Oktay, O., Schlemper, J., Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning where to look for the pancreas. In: Med. Imag. Deep Learn. (2018)
18. Petitjean, C., Dacher, J.N.: A review of segmentation methods in short axis cardiac MR images. Med. Imag. Anal. **15**(2), 169–84 (2011)
19. Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M.: BASNet: Boundary-aware salient object detection. In: IEEE Conf. Comp. Vis. Pattern Recog. pp. 7471–81 (2019)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Med. Imag. Comput. Comp.-Assist. Interv. pp. 234–41 (2015)
21. Royer, L., Richmond, D., Rother, C., Andres, B., Kainmueller, D.: Convexity shape constraints for image segmentation. In: IEEE Comp. Vis. Pattern Recog. pp. 402–10 (2016)
22. Sanner, A., Gonzalez, C., Mukhopadhyay, A.: How reliable are out-of-distribution generalization methods for medical image segmentation? In: DAGM German Conf. Pattern Recog. pp. 604–17 (2021)
23. Shaaf, Z., Jamil, M., Ambar, R., Alattab, A., Yahya, A., Asiri, Y.: Automatic left ventricle segmentation from short-axis cardiac MRI images based on fully convolutional neural network. Diagnostics (Basel) **12**(2), 414 (2022)
24. Shigwan, S., Gaikwad, A., Awate, S.: 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. Symp. Biomed. Imaging. pp. 1149–53 (2020)
25. Shit, S., Paetzold, J., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J., Bauer, U., Menze, B.: clDice - a novel topology-preserving loss function for tubular structure segmentation. In: IEEE Comp. Vis. Pattern Recog. pp. 16555–64 (2021)
26. Smith, T., Zhang, S., Erkanli, A., Frush, D., Samei, E.: Variability in image quality and radiation dose within and across 97 medical facilities. J. Med. Imag. **8**, 52105 (2021)
27. Suinesiaputra, A., Cowan, B.R., Al-Agamy, A.O., Elattar, M.A., Ayache, N., Fahmy, A.S., Khalifa, A.M., Medrano-Gracia, P., Jolly, M.P., Kadish, A.H., et al.:

A collaborative resource to build consensus for automated left ventricular segmentation of cardiac mr images. Medical image analysis **18**(1), 50–62 (2014)
28. Sun, F., Luo, Z., Li, S.: Boundary difference over union loss for medical image segmentation. In: MICCAI 2023. vol. 14223, pp. 292–301 (2023)
29. Tasman, W., Jaeger, E.: Duane's Ophthalmology. Lippincott Williams (2013)
30. Varma, H., Gaikwad, A.V., Awate, S.P.: Adversarial training with multiscale boundary-prediction dnn for robust topologically-constrained segmentation in ood images. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2023)
31. Wang, B., Gu, X., Fan, C., Xie, H., Zhang, S., Tian, X., Gu, L.: Sparse group composition for robust left ventricular epicardium segmentation. Computerized Medical Imaging and Graphics **46**, 56–63 (2015)
32. Wang, L., Gu, J., Chen, Y., Liang, Y., Zhang, W., Pu, J., Chen, H.: Automated segmentation of the optic disc from fundus images using an asymmetric deep learning network. Pattern recognition **112** (2021)
33. Wu, J., Wang, K., Shang, Z., Xu, J., Ding, D., Li, X., Yang, G.: Oval shape constraint based optic disc and cup segmentation in fundus photographs. In: Brit. Mach. Vis. Conf. (2019)
34. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual U-Net. IEEE Geosc. Rem. Send. Letters **15**(5), 749–53 (2018)
35. Zhang, Z., Yin, F., Liu, J., Wong, W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: Origa-light: An online retinal fundus image database for glaucoma analysis and research. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology pp. 3065–68 (2010)
36. Zhuo, J., Gullapalli, R.: AAPM/RSNA physics tutorial for residents: MR artifacts, safety, and quality control. Radiographics **26**(1), 275–97 (2006)