MIDL
Lübeck 2021

Beyond pixel-wise supervision:

Semantic segmentation with higher-order shape descriptors

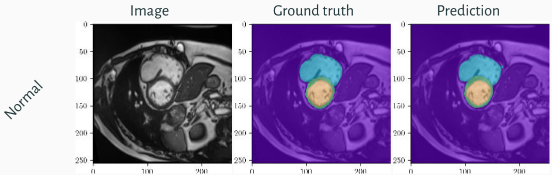
Hoel Kervadec[1,2] Houda Bahig[2] Laurent Letourneau-Guillon[2]
Jose Dolz[1] Ismail Ben Ayed[1,2]

[1]ÉTS Montréal, [2]CRChum

Medical Imaging with Deep Learning — 7-9 July 2021

Motivation

Standard losses treat segmentation as independent classification



Standard losses treat segmentation as independent classification

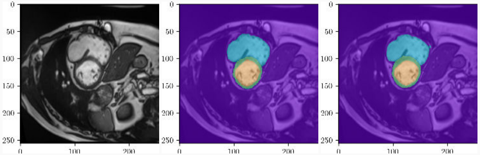
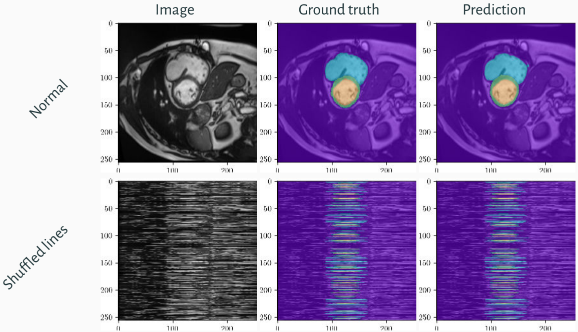


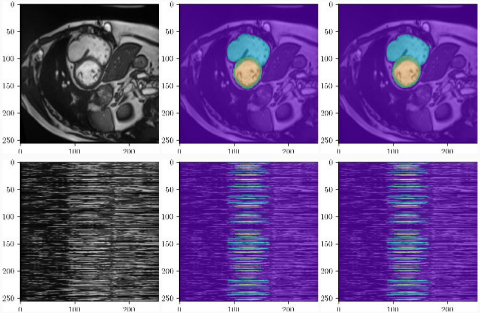
	Image	Ground truth	Prediction	Cross-entropy	Dice
Normal				0.09	0.92

Standard losses treat segmentation as independent classification



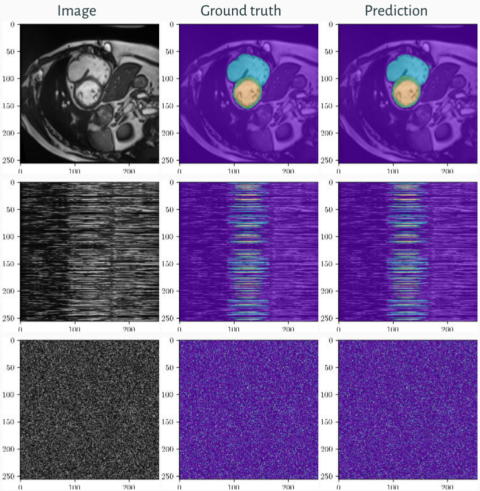
Cross-entropy	Dice
0.09	0.92

Standard losses treat segmentation as independent classification



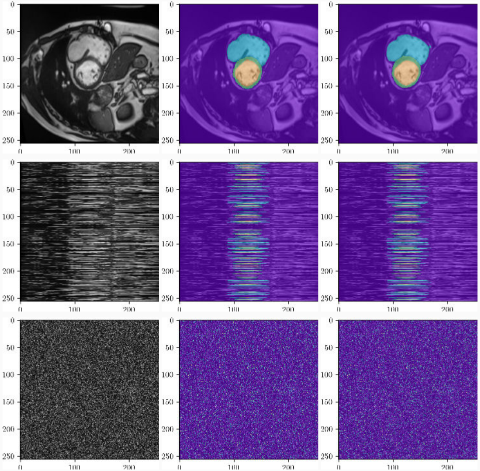
	Cross-entropy	Dice
Normal	0.09	0.92
Shuffled lines	0.09	0.92

Standard losses treat segmentation as independent classification



	Cross-entropy	Dice
Normal	0.09	0.92
Shuffled lines	0.09	0.92
Shuffled pixels	0.09	0.92

Standard losses treat segmentation as independent classification



	Cross-entropy	Dice	Centroid	Distance to centroid
Normal	0.09	0.92		
Shuffled lines	0.09	0.92		
Shuffled pixels	0.09	0.92		

Standard losses treat segmentation as independent classification

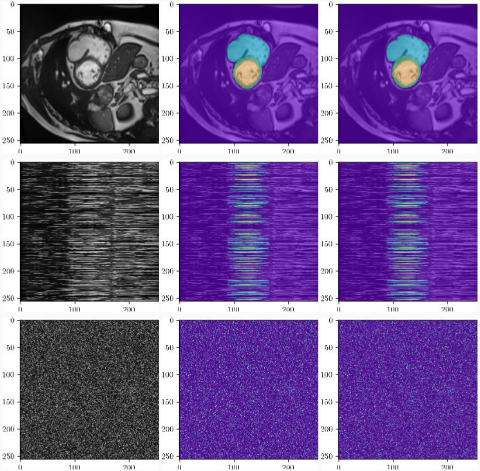
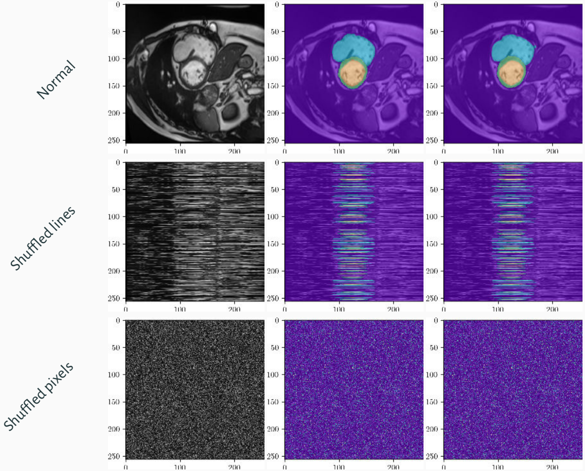


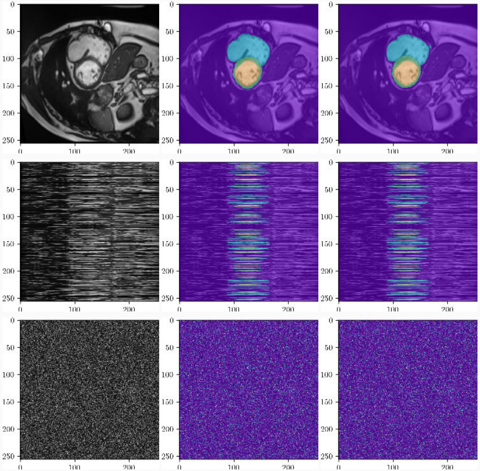
	Image / Ground truth / Prediction	Cross-entropy	Dice	Centroid	Distance to centroid
Normal		0.09	0.92	C=1 (83,127) C=2 (121,123) C=3 (126,125)	C=1 (12,22) C=2 (18,16) C=3 (11,11)
Shuffled lines		0.09	0.92		
Shuffled pixels		0.09	0.92		

Standard losses treat segmentation as independent classification



	Cross-entropy	Dice	Centroid	Distance to centroid
Normal	0.09	0.92	C=1 (83,127) C=2 (121,123) C=3 (126,125)	C=1 (12,22) C=2 (18,16) C=3 (11,11)
Shuffled lines	0.09	0.92	C=1 (132,127) C=2 (121,123) C=3 (123,125)	C=1 (75,22) C=2 (78,16) C=3 (70,11)
Shuffled pixels	0.09	0.92		

Standard losses treat segmentation as independent classification



	Cross-entropy	Dice	Centroid	Distance to centroid
Normal	0.09	0.92	C=1 (83,127) C=2 (121,123) C=3 (126,125)	C=1 (12,22) C=2 (18,16) C=3 (11,11)
Shuffled lines	0.09	0.92	C=1 (132,127) C=2 (121,123) C=3 (123,125)	C=1 (75,22) C=2 (78,16) C=3 (70,11)
Shuffled pixels	0.09	0.92	C=1 (129,131) C=2 (130,129) C=3 (129,131)	C=1 (74,74) C=2 (73,73) C=3 (74,74)

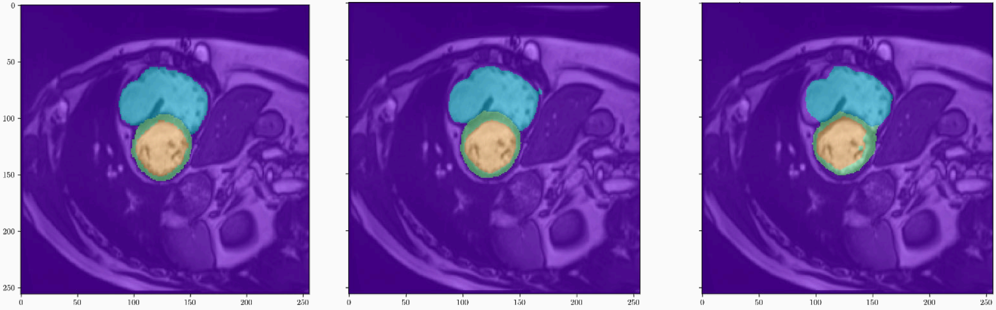
Contributions

- Re-introduce shape descriptors, in the context of deep learning-based semantic segmentation;

Contributions

- Re-introduce shape descriptors, in the context of deep learning-based semantic segmentation;
- propose a way to supervise using *only* shape descriptors.

Ground truth	Trained with cross-entropy 65536 discrete labels	Trained with shape descriptors 4 descriptors per class



Computing shape descriptors

Shape moments

General *shape* moment:

$$\mu_{p,q}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} s_{\boldsymbol{\theta}}{}^{(i,k)} x_{(i)}^{p} y_{(i)}^{q},$$

Shape moments

General *shape* moment:

$$\mu_{p,q}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} {s_{\boldsymbol{\theta}}}^{(i,k)} x_{(i)}^{p} y_{(i)}^{q},$$

p, q moments order;

Shape moments

General *shape* moment:

$$\mu^{(k)}_{p,q}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} {s_{\boldsymbol{\theta}}}^{(i,k)} x^p_{(i)} y^q_{(i)},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;

Shape moments

General *shape* moment:

$$\mu_{p,q}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} {s_{\boldsymbol{\theta}}}^{(i,k)} x_{(i)}^{p} y_{(i)}^{q},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;

Shape moments

General *shape* moment:

$$\mu^{(k)}_{p,q}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} s_{\boldsymbol{\theta}}{}^{(i,k)} x^{p}_{(i)} y^{q}_{(i)},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;

Shape moments

General *shape* moment:

$$\mu_{p,q}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} {s_{\boldsymbol{\theta}}}^{(i,k)} x_{(i)}^{p} y_{(i)}^{q},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;
x, y	pixels-coordinates (2D).

Shape moments

General *shape* moment:

$$\mu_{p,q}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} {s_{\boldsymbol{\theta}}}^{(i,k)} x_{(i)}^{p} y_{(i)}^{q},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;
x, y	pixels-coordinates (2D).

Volume:

$$\mathfrak{V}^{(k)}(s_{\boldsymbol{\theta}}) := \mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}})$$

Shape moments

General *shape* moment:

$$\mu^{(k)}_{p,q}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} {s_{\boldsymbol{\theta}}}^{(i,k)} x^p_{(i)} y^q_{(i)},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;
x, y	pixels-coordinates (2D).

Volume:

$$\mathfrak{V}^{(k)}(s_{\boldsymbol{\theta}}) := \mu^{(k)}_{0,0}(s_{\boldsymbol{\theta}}) = \sum_{i \in \Omega} s^{(i,k)}_{\boldsymbol{\theta}}.$$

Shape moments

General *shape* moment:

$$\mu_{p,q}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} s_{\boldsymbol{\theta}}{}^{(i,k)} x_{(i)}^{p} y_{(i)}^{q},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;
x, y	pixels-coordinates (2D).

Volume:

$$\mathfrak{V}^{(k)}(s_{\boldsymbol{\theta}}) := \mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}}) = \sum_{i \in \Omega} s_{\boldsymbol{\theta}}^{(i,k)}.$$

Centroid: “average” of pixels coordinates

$$\mathfrak{C}^{(k)}(s_{\boldsymbol{\theta}}) := \left(\frac{\mu_{1,0}^{(k)}(s_{\boldsymbol{\theta}})}{\mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}})}, \frac{\mu_{0,1}^{(k)}(s_{\boldsymbol{\theta}})}{\mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}})} \right)$$

Shape moments

General *shape* moment:

$$\mu_{p,q}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i \in \Omega} {s_{\boldsymbol{\theta}}}^{(i,k)} x_{(i)}^{p} y_{(i)}^{q},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;
x, y	pixels-coordinates (2D).

Volume:

$$\mathfrak{V}^{(k)}(s_{\boldsymbol{\theta}}) := \mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}}) = \sum_{i \in \Omega} s_{\boldsymbol{\theta}}^{(i,k)}.$$

Centroid: "average" of pixels coordinates

$$\mathfrak{C}^{(k)}(s_{\boldsymbol{\theta}}) := \left(\frac{\mu_{1,0}^{(k)}(s_{\boldsymbol{\theta}})}{\mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}})}, \frac{\mu_{0,1}^{(k)}(s_{\boldsymbol{\theta}})}{\mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}})} \right) = \left(\frac{\mu_{1,0}^{(k)}(s_{\boldsymbol{\theta}})}{\mathfrak{V}^{(k)}}, \frac{\mu_{0,1}^{(k)}(s_{\boldsymbol{\theta}})}{\mathfrak{V}^{(k)}} \right).$$

Central moment

General *central* moment:

$$\bar{\mu}^{(k)}_{p,q} := \sum_{i \in \Omega} s^{(i,k)}_{\boldsymbol{\theta}} \left(x_{(i)} - \frac{\mu^{(k)}_{1,0}}{\mu^{(k)}_{0,0}} \right)^{p} \left(y_{(i)} - \frac{\mu^{(k)}_{0,1}}{\mu^{(k)}_{0,0}} \right)^{q},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;
x, y	pixel coordinates (2D).

Central moment

General *central* moment:

$$\bar{\mu}_{p,q}^{(k)} := \sum_{i \in \Omega} s_{\boldsymbol{\theta}}^{(i,k)} \left(x_{(i)} - \frac{\mu_{1,0}^{(k)}}{\mu_{0,0}^{(k)}} \right)^{p} \left(y_{(i)} - \frac{\mu_{0,1}^{(k)}}{\mu_{0,0}^{(k)}} \right)^{q},$$

p, q	moments order;
$s_{\boldsymbol{\theta}}$	softmax predictions;
k	class;
$i \in \Omega$	pixels in the image;
x, y	pixel coordinates (2D).

Average distance to the centroid: "standard deviation" of pixels coordinates

$$\mathfrak{D}^{(k)}(s_{\boldsymbol{\theta}}) := \left(\sqrt[2]{\frac{\bar{\mu}_{2,0}^{(k)}(s_{\boldsymbol{\theta}})}{\mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}})}}, \sqrt[2]{\frac{\bar{\mu}_{0,2}^{(k)}(s_{\boldsymbol{\theta}})}{\mu_{0,0}^{(k)}(s_{\boldsymbol{\theta}})}} \right).$$

Other descriptors

Length: "perimeter" of the prediction

$$\mathfrak{L}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i,j \in \mathcal{G}_\Omega} |s_{\boldsymbol{\theta}}^{(i,k)} - s_{\boldsymbol{\theta}}^{(j,k)}|, \qquad i, j \in \mathcal{G}_\Omega \quad \text{pairs of neighbor pixels.}$$

Other descriptors

Length: "perimeter" of the prediction

$$\mathfrak{L}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i,j \in \mathcal{G}_\Omega} |s_{\boldsymbol{\theta}}^{(i,k)} - s_{\boldsymbol{\theta}}^{(j,k)}|, \qquad i, j \in \mathcal{G}_\Omega \quad \text{pairs of neighbor pixels.}$$

Ratio of descriptors:

$$\mathfrak{R}_{\mathfrak{f}}^{(k,l)}(s_{\boldsymbol{\theta}}) := \frac{\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}})}{\mathfrak{f}^{(l)}(s_{\boldsymbol{\theta}})} \qquad \mathfrak{f} \in \{\mathfrak{V}, \mathfrak{C}, \mathfrak{D}, \mathfrak{L}\}$$

Other descriptors

Length: "perimeter" of the prediction

$$\mathfrak{L}^{(k)}(s_{\boldsymbol{\theta}}) := \sum_{i,j \in \mathcal{G}_\Omega} |s_{\boldsymbol{\theta}}^{(i,k)} - s_{\boldsymbol{\theta}}^{(j,k)}|,$$

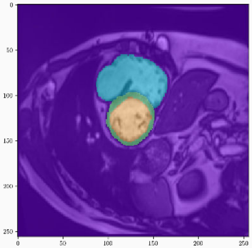
$i, j \in \mathcal{G}_\Omega$ pairs of neighbor pixels.

Ratio of descriptors:

$$\mathfrak{R}_{\mathfrak{f}}^{(k,l)}(s_{\boldsymbol{\theta}}) := \frac{\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}})}{\mathfrak{f}^{(l)}(s_{\boldsymbol{\theta}})}$$

$\mathfrak{f} \in \{\mathfrak{V}, \mathfrak{C}, \mathfrak{D}, \mathfrak{L}\}$

Example:



The myocardium (green) surrounds the left-ventricle (yellow):

$$\frac{\mathfrak{L}^{(\text{green})}}{\mathfrak{L}^{(\text{yellow})}} \geq 2.$$

Training with descriptors

Series of losses optimized with SGD:

$$\underset{\boldsymbol{\theta}}{\arg\min} \sum_{\mathfrak{f}\in\{\mathfrak{V},\mathfrak{C},\mathfrak{D},\mathfrak{L}\}} \sum_{k} \left[\widetilde{\psi}_t \left(0.9{\tau_{\mathfrak{f}}}^{(k)} - \mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) \right) + \widetilde{\psi}_t \left(\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) - 1.1{\tau_{\mathfrak{f}}}^{(k)} \right) \right],$$

$\tau_{\mathfrak{f}}$: ground-truth descriptor, $\mathfrak{f}(s_{\boldsymbol{\theta}})$: descriptor on prediction, $\widetilde{\psi}_t(z)$: extended log-barrier:

Training with descriptors

Series of losses optimized with SGD:

$$\underset{\boldsymbol{\theta}}{\arg\min} \sum_{\mathfrak{f}\in\{\mathfrak{V},\mathfrak{C},\mathfrak{D},\mathfrak{L}\}} \sum_{k} \left[\widetilde{\psi}_t\left(0.9\tau_{\mathfrak{f}}{}^{(k)} - \mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}})\right) + \widetilde{\psi}_t\left(\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) - 1.1\tau_{\mathfrak{f}}{}^{(k)}\right)\right],$$

$\tau_{\mathfrak{f}}$: ground-truth descriptor, $\mathfrak{f}(s_{\boldsymbol{\theta}})$: descriptor on prediction, $\widetilde{\psi}_t(z)$: extended log-barrier:

Training with descriptors

Series of losses optimized with SGD:

$$\arg\min_{\boldsymbol{\theta}} \sum_{\mathfrak{f}\in\{\mathfrak{V},\mathfrak{C},\mathfrak{D},\mathfrak{L}\}} \sum_{k} \left[\widetilde{\psi}_t \left(0.9\tau_{\mathfrak{f}}{}^{(k)} - \mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}})\right) + \widetilde{\psi}_t \left(\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) - 1.1\tau_{\mathfrak{f}}{}^{(k)}\right)\right],$$

$\tau_{\mathfrak{f}}$: ground-truth descriptor, $\mathfrak{f}(s_{\boldsymbol{\theta}})$: descriptor on prediction, $\widetilde{\psi}_t(z)$: extended log-barrier:

Training with descriptors

Series of losses optimized with SGD:

$$\arg\min_{\boldsymbol{\theta}} \sum_{\mathfrak{f}\in\{\mathfrak{V},\mathfrak{C},\mathfrak{D},\mathfrak{L}\}} \sum_{k} \left[\widetilde{\psi}_t\left(0.9{\tau_{\mathfrak{f}}}^{(k)} - \mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}})\right) + \widetilde{\psi}_t\left(\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) - 1.1{\tau_{\mathfrak{f}}}^{(k)}\right)\right],$$

$\tau_{\mathfrak{f}}$: ground-truth descriptor, $\mathfrak{f}(s_{\boldsymbol{\theta}})$: descriptor on prediction, $\widetilde{\psi}_t(z)$: extended log-barrier:

Training with descriptors

Series of losses optimized with SGD:

$$\arg\min_{\boldsymbol{\theta}} \sum_{\mathfrak{f}\in\{\mathfrak{V},\mathfrak{C},\mathfrak{D},\mathfrak{L}\}} \sum_{k} \left[\widetilde{\psi}_t\left(0.9{\tau_{\mathfrak{f}}}^{(k)} - \mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}})\right) + \widetilde{\psi}_t\left(\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) - 1.1{\tau_{\mathfrak{f}}}^{(k)}\right)\right],$$

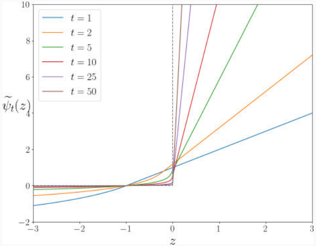
$\tau_{\mathfrak{f}}$: ground-truth descriptor, $\mathfrak{f}(s_{\boldsymbol{\theta}})$: descriptor on prediction, $\widetilde{\psi}_t(z)$: extended log-barrier:

Training with descriptors

Series of losses optimized with SGD:

$$\arg\min_{\boldsymbol{\theta}} \sum_{\mathfrak{f}\in\{\mathfrak{V},\mathfrak{C},\mathfrak{D},\mathfrak{L}\}} \sum_{k} \left[\widetilde{\psi}_t \left(0.9\tau_{\mathfrak{f}}^{(k)} - \mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) \right) + \widetilde{\psi}_t \left(\mathfrak{f}^{(k)}(s_{\boldsymbol{\theta}}) - 1.1\tau_{\mathfrak{f}}^{(k)} \right) \right],$$

$\tau_{\mathfrak{f}}$: ground-truth descriptor, $\mathfrak{f}(s_{\boldsymbol{\theta}})$: descriptor on prediction, $\widetilde{\psi}_t(z)$: extended log-barrier:

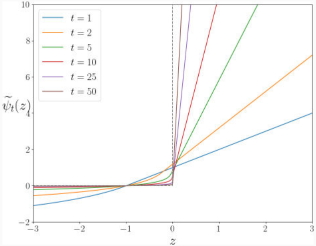


Training with descriptors

Series of losses optimized with SGD:

$$\arg\min_{\boldsymbol{\theta}} \sum_{\mathfrak{f} \in \{\mathfrak{V},\mathfrak{C},\mathfrak{D},\mathfrak{L}\}} \sum_{\mathrm{k}} \left[\widetilde{\psi}_{\mathrm{t}} \left(0.9{\tau_{\mathfrak{f}}}^{(\mathrm{k})} - \mathfrak{f}^{(\mathrm{k})}(\mathrm{s}_{\boldsymbol{\theta}}) \right) + \widetilde{\psi}_{\mathrm{t}} \left(\mathfrak{f}^{(\mathrm{k})}(\mathrm{s}_{\boldsymbol{\theta}}) - 1.1{\tau_{\mathfrak{f}}}^{(\mathrm{k})} \right) \right],$$

$\tau_{\mathfrak{f}}$: ground-truth descriptor, $\mathfrak{f}(\mathrm{s}_{\boldsymbol{\theta}})$: descriptor on prediction, $\widetilde{\psi}_{\mathrm{t}}(\mathrm{z})$: extended log-barrier:



Allows us to enforce $0.9\tau_{\mathfrak{f}} \leq \mathfrak{f}(\mathrm{s}_{\boldsymbol{\theta}}) \leq 1.1\tau_{\mathfrak{f}}$.

Experiments

Experimental setting

Test on two datasets: ACDC (heart segmentation, 4 classes) and PROMISE12 (prostate, 2 classes).

Experimental setting

Test on two datasets: ACDC (heart segmentation, 4 classes) and PROMISE12 (prostate, 2 classes).

Supervise *only* with descriptors: volume, centroid, distance to centroid and length.

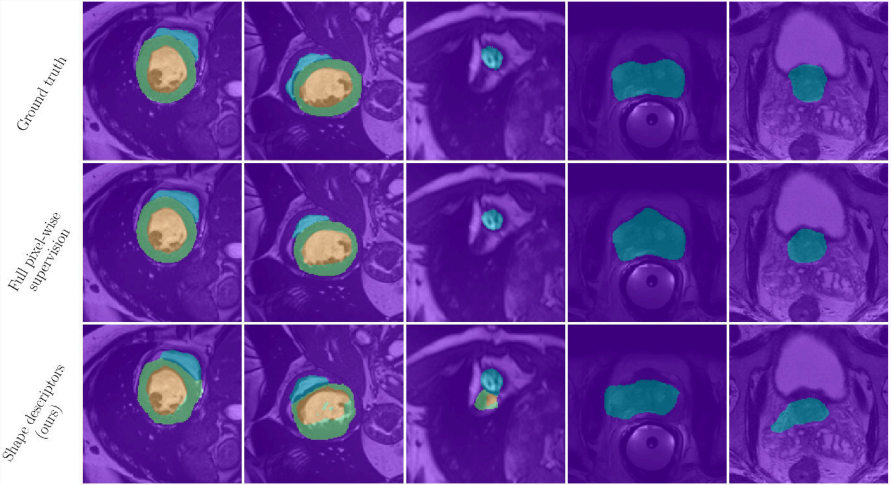
Experimental setting

Test on two datasets: ACDC (heart segmentation, 4 classes) and PROMISE12 (prostate, 2 classes).

Supervise *only* with descriptors: volume, centroid, distance to centroid and length.

Compare to pixel-wise cross-entropy.

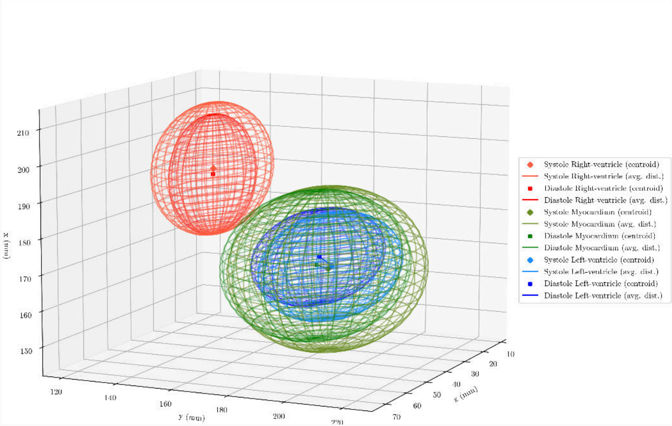
Results



Metrics numbers available in the paper.

Future works: generalization to 3D and time series

Descriptors as bounds can be valid for multiple time-points:



Ellipsoids for each class at two time points: systole and diastole (ACDC dataset).

Conclusion

Different way to think about image segmentation;

Conclusion

Different way to think about image segmentation;

many more descriptors can be defined;

Conclusion

Different way to think about image segmentation;

many more descriptors can be defined;

can be a way to encode anatomical priors;

Conclusion

Different way to think about image segmentation;

many more descriptors can be defined;

can be a way to encode anatomical priors;

opens new ways to use existing annotations;

Conclusion

Different way to think about image segmentation;

many more descriptors can be defined;

can be a way to encode anatomical priors;

opens new ways to use existing annotations;

code available online: `https://github.com/HKervadec/shape_descriptors`.