ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Progressive Adversarial Semantic Segmentation

Abdullah-Al-Zubaer Imran[1], Demetri Terzopoulos[1,2]
[1]UCLA Computer Science Department, Los Angeles, CA, USA
[2]VoxelCloud, Inc., Los Angeles, CA, USA
Correspondence: aimran@Stanford.edu | http://web.stanford.edu/~aimran/

Highlights

- **Goal:** Domain generalization for medical image segmentation from single source training data, in presence of large domain shifts.
- **Solution:** Enlarging the scope of train data distributions through transformation of the given examples and employing consistency regularization against predictions on the non-transformed inputs.
- **Contribution:** Progressive Adversarial Semantic Segmentation (PASS) for improved and consistent pixel-wise segmentation predictions without requiring any domain-specific data.

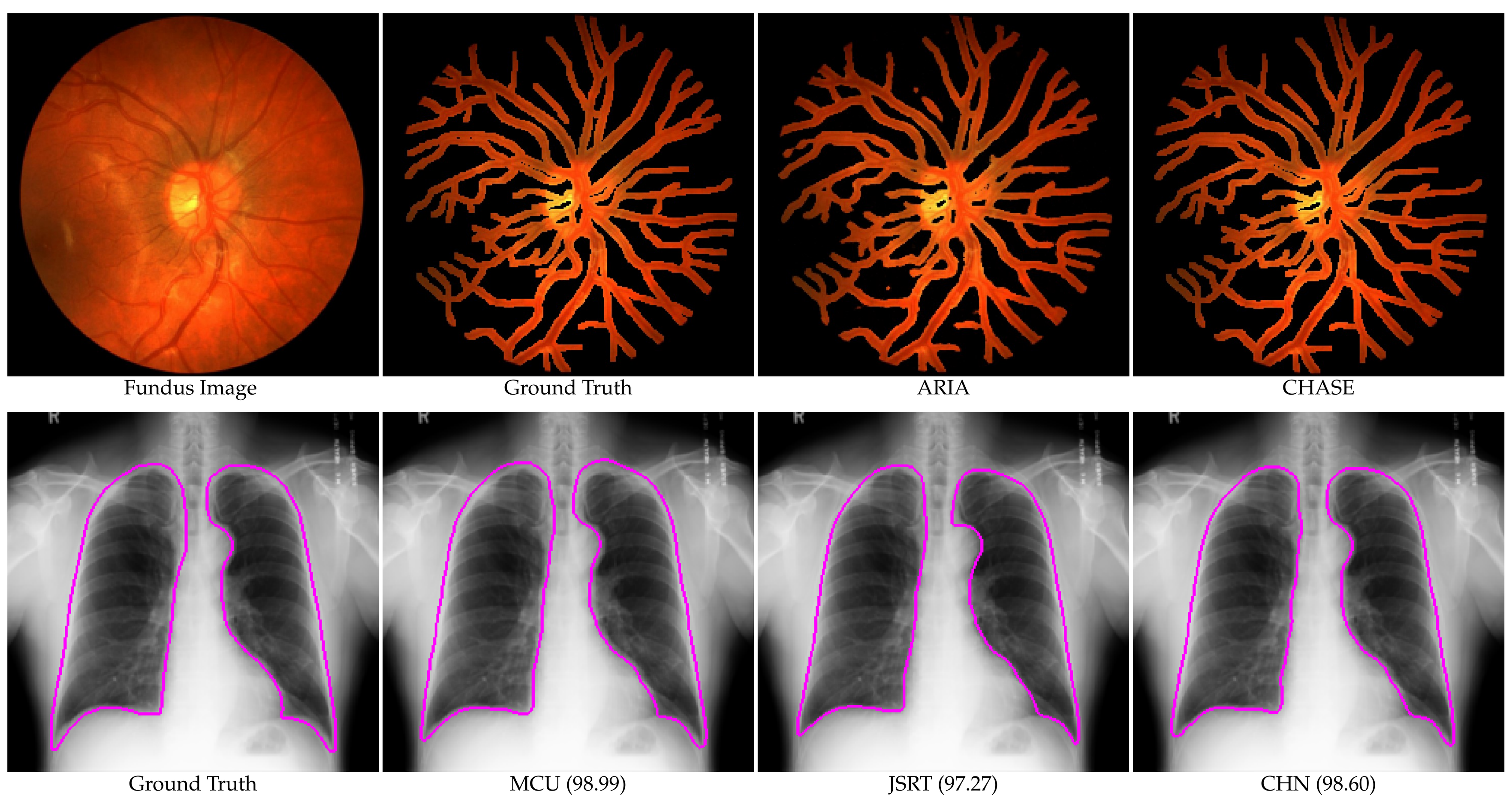

Figure 1: Consistency of the in-domain and cross-domain segmentation predictions by our PASS model: (Top) Visualization of retinal vessel segmentation from a fundus image when trained on the ARIA and CHASE datasets. (Bottom) Segmentation of a chest X-ray from the MCU dataset when the model is trained on the MCU, JSRT, and CHN datasets.

Methods

- A progressive U-Net with some careful adjustments plus side-adversary and side-supervision capabilities at different resolutions of the segmentation predictions.
- A shape encoder matches the latent representation of the stacked input and segmentation output with the stacked input and reference so that the model becomes shape-aware while mapping an input to the segmentation mask.

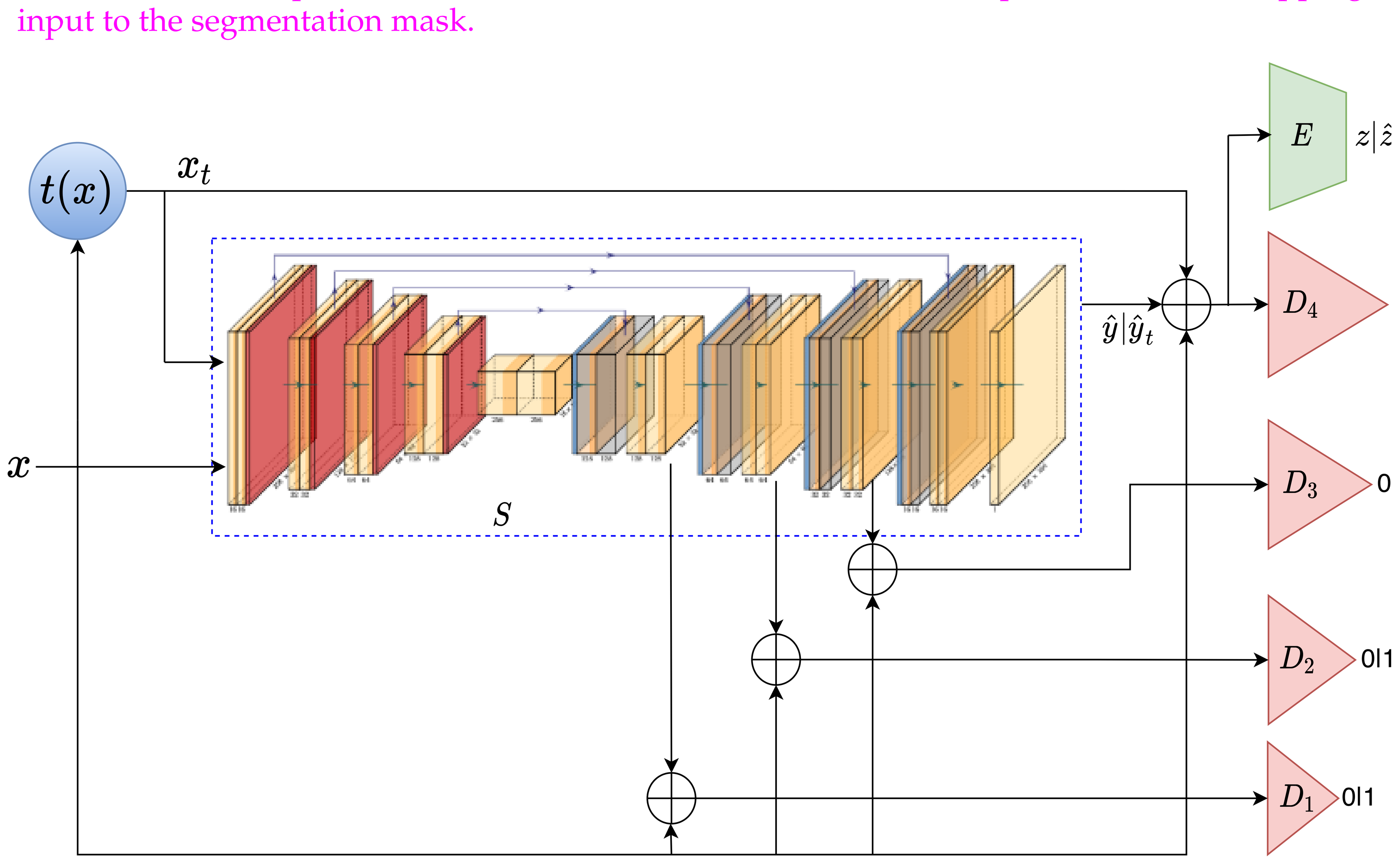

Figure 2: Schematic of the PASS model. The segmentation mask generator S takes either input x or its transformed version $x_t = t(x)$. The generated side outputs are passed to the corresponding discriminators, D_1, D_2, and D_3, and the final outputs y_t or $\hat{y}$ are passed to the discriminator D_4. The shape encoder E also takes y or $\hat{y}$ concatenated with x as input to yield the latent vector z or $\hat{z}$.

S Loss	D Loss	E Loss
$L_S^{\text{seg}} = \sum_{k=1}^{K} w_k L_S^{\text{Dice}}(y_k, \hat{y}_k)$		
$L_S^{\text{seg}}(\hat{y}, \hat{y}_t) = \sum_{i=1}^{N} \lvert(\hat{y}(i) - \hat{y}_t(i)) \log(\hat{y}(i)/\hat{y}_t(i))\rvert$	$L_D^{\text{adv}}(x, y) = \sum_{k=1}^{K} -\mathbb{E}_{x_k, y_k} \log[1 - D_k(x_k, y_k)]$	$L_E^{\text{lat}}(x, y, \hat{y}) = \lvert\lvert z - \hat{z}\rvert\rvert$
$L_S^{\text{feat}}(x, y, \hat{y}) = \sum_{k=1}^{K} \lvert\lvert f(x_k, y_k) - f(x_k, \hat{y}_k)\rvert\rvert^2$	$L_D^{\text{adv}}(x, \hat{y}) = \sum_{k=1}^{K} -\mathbb{E}_{x_k, \hat{y}_k} \log[D_k(x_k, \hat{y}_k)]$	
$L_S^{\text{adv}}(x, \hat{y}) = \sum_{k=1}^{K} -\mathbb{E}_{x_k, \hat{y}_k} \log[1 - D_k(x_k, \hat{y}_k)]$		

Table 1: Learning objectives for the three networks in the proposed PASS model.

Experiment Details

Dataset	#Fundus images			Data splits			Dataset	#X-ray images			Data splits		
	Total	Healthy	Diseased	Train	Val	Test		Total	Healthy	Diseased	Train	Val	Test
DRIVE	40	33	7	18	2	20	MCU	138	80	58	93	10	35
STARE	20	10	10	10	2	8	JSRT	247	93	154	111	13	123
CHASE	28	20	8	17	5	6	CHN	566	279	287	381	43	142
ARIA	143	61	82	121	5	17							
HRF	45	15	30	26	5	14							

Table 2: Partitioning of the fundoscopic and chest X-ray datasets used in our experiments.

- **Inputs**: All the images were resized and normalized to $256 \times 256 \times 3$ for the color fundus images and $256 \times 256 \times 1$ for the monochrome chest X-rays before feeding them to the network.
- **Networks**: The segmentor S utilizes 3×3 convolutions with the following feature map sizes: 16, 16, 32, 32, 64, 64, 128, 128 in the contraction; 256, 256 in the bottleneck; 128, 128, 64, 64, 32, 32, 16, 16 in the expansion; and a single-channel convolution to obtain the final output.
- **Transformation**: Within $t(x)$ for every x, left-right flip was performed with 0.5 probability, sharpen by no to full effect, rotate by $-90°$ to $90°$, and shear by $-8°$ to $8°$.

Results

- **Vasculature Segmentation**: In vessel segmentation, PASS achieved an overall average Dice score of 85.30 and cross-domain score of 83.74 (domain gap of 8.32), when trained on the small CHASE dataset.
- When trained on the relatively larger ARIA dataset, the overall average Dice score of 88.72 and cross-domain score of 87.88 were achieved with domain gap of 4.22.

Model	Train on →	CHASE						ARIA					
	Test on →	CHASE	DRIVE	ARIA	STARE	HRF	Avg	CHASE	DRIVE	ARIA	STARE	HRF	Avg
U-Net		80.40	63.20	64.50	66.76	63.82	67.74	76.70	77.30	72.00	71.28	72.30	73.90
U-Net+CRF		81.20	65.40	62.60	56.40	63.60	65.80	78.40	69.50	73.00	64.60	73.50	71.80
PU-Net		81.58	64.04	63.03	66.20	62.66	67.50	76.7	77.3	72.0	71.28	72.3	73.9
AttnU-Net		81.37	65.23	62.91	64.28	65.72	67.90	76.7	77.3	72.0	71.28	72.3	73.9
ProgU-Net		82.91	61.02	63.28	66.58	63.43	67.44	47.21	64.54	70.56	66.57	60.17	61.81
ProgU-NetSS		80.16	62.13	62.41	65.44	63.78	66.78	57.96	65.93	74.47	69.08	60.08	65.50
V-GAN		79.70	71.50	64.20	61.00	66.40	68.50	68.70	75.80	69.90	66.20	69.30	70.00
AU-Net		82.20	63.20	61.84	67.17	63.37	67.56	68.06	70.21	78.12	74.95	69.69	72.21
APPU-Net		82.58	62.50	61.22	66.17	62.60	67.01	66.20	69.68	78.48	76.34	69.31	72.00
UDA		72.30	69.30	68.20	64.70	67.40	68.40	71.50	72.90	73.20	71.30	70.70	71.90
ErrorNet		81.50	73.20	66.50	65.20	68.60	71.00	76.70	78.90	72.00	74.00	72.60	74.80
PASS without $t(x)$		89.06	80.76	80.72	82.72	75.26	81.70	85.06	88.14	91.92	90.78	82.62	87.70
PASS		**91.96**	**84.96**	**84.18**	**86.84**	**78.57**	**85.30**	**86.32**	**90.55**	**92.08**	**91.50**	**83.15**	**88.72**

Table 3: Comparison between PASS and other performance baselines for retinal vessel segmentation.

- **Pulmonary Segmentation**: With PASS, we have a domain gap of only 1.32 when trained on MCU, 2.39 when trained on JSRT, and 0.44 when trained on CHN.
- The poorer performance of the PASS model without $t(x)$ justifies its inclusion and the logit-wise distribution matching as the consistency regularization.

Model	Train on →	MCU				JSRT				CHN			
	Test on →	MCU	JSRT	CHN	Avg	MCU	JSRT	CHN	Avg	MCU	JSRT	CHN	Avg
U-Net		97.67	39.39	94.48	77.18	92.00	95.02	90.54	92.58	93.72	43.46	95.84	77.67
PU-Net		97.89	21.24	**97.84**	72.33	84.97	94.94	73.68	84.53	93.57	73.88	95.90	87.78
AttnU-Net		97.86	30.31	94.07	74.08	6.70	94.95	65.00	55.55	81.25	74.24	95.56	83.68
ProgU-Net		97.83	10.98	91.32	66.71	34.89	95.20	86.28	72.12	84.79	60.03	95.35	80.06
ProgU-NetSS		97.90	33.98	95.32	75.33	13.16	95.09	65.00	57.75	94.24	67.29	95.63	85.72
AU-Net		97.86	94.68	95.08	95.87	89.12	97.85	92.46	93.14	95.58	95.88	96.22	95.89
APPU-Net		97.81	95.07	94.77	95.88	90.46	97.80	91.76	93.34	95.72	96.25	96.11	96.03
CyUDA		95.61	92.84	–	94.23	–	–	–	–	–	–	–	–
SeUDA		95.61	94.51	–	– –	–	–	–	–	–	–	–	
CoDAGAN		–	–	–	–	84.58	96.45	88.99	90.01	–	–	–	–
PASS without $t(x)$		97.74	96.43	96.76	96.98	95.11	98.26	95.92	96.43	96.62	96.11	97.61	96.68
PASS		**98.22**	**96.56**	97.24	**97.34**	**95.70**	**98.27**	**96.06**	**96.68**	**97.27**	**97.15**	**97.65**	**97.36**

Table 4: Comparison between PASS and other performance baselines for pulmonary segmentation.

Conclusions

- PASS, an innovative semantic segmentation model mitigates the domain shift problem in learning from small annotated training datasets.
- Future work will focus on evaluating PASS on other image segmentation tasks as well as assessing its effectiveness in iterative and active learning settings.