I. Training process

Training of the DAFNet can be described with the pseudocode described in Algorithm 1. This corresponds to a case with fully annotated data from two input modalities. If no labels are available for the second modality, then lines 64 and 71 are omitted.

Algorithm 1 Pseudocode for training DAFNet with fully annotated images from two modalities.

1: {parameter initialisation}
2: $\theta_{a1} \leftarrow$ parameters of encoder $f_{anatomy}$ for modality 1
3: $\theta_{a2} \leftarrow$ parameters of encoder $f_{anatomy}$ for modality 2
4: $\theta_m \leftarrow$ parameters of modality encoder $f_{modality}$
5: $\theta_h \leftarrow$ parameters of segmentor h
6: $\theta_g \leftarrow$ parameters of decoder g
7: $\theta_{stn} \leftarrow$ parameters of spatial transformer stn
8: $\psi \leftarrow$ parameters of mask discriminator D_M
9: $\omega_1 \leftarrow$ weights of image discriminator $D_{I,1}$ (modality 1)
10: $\omega_2 \leftarrow$ weights of image discriminator $D_{I,2}$ (modality 2)
11: **repeat**
12: {forward pass for modality 1 inputs}
13: $x_1, l_1 \leftarrow$ get batch of images and masks of modality 1
14: $s_1 \leftarrow f_{anatomy}(x_1)$
15: $z_1^{mean}, z_1^{std} \leftarrow f_{modality}(x_1, s_1)$
16: $\epsilon \leftarrow$ random sample from $\mathcal{N}(\mathbf{0}, \mathbf{I})$
17: $z_1 \leftarrow z_1^{mean} + z_1^{std} \odot \epsilon$ {reparameterisation trick}
18: $m_1 \leftarrow h(s_1)$
19: $y_1 \leftarrow g(s_1, z_1)$
20: $z_{sample1} \leftarrow$ random sample from $\mathcal{N}(\mathbf{0}, \mathbf{I})$
21: $z_{rec1} \leftarrow f_{modality}(g(s_1, z_{sample1}), x_1)$
22:
23: {forward pass for modality 2 inputs}
24: $x_2, l_2 \leftarrow$ get batch of images and masks of modality 2
25: $s_2 \leftarrow f_{anatomy}(x_2)$
26: $z_2^{mean}, z_2^{std} \leftarrow f_{modality}(x_2, s_2)$
27: $\epsilon \leftarrow$ random sample from $\mathcal{N}(\mathbf{0}, \mathbf{I})$
28: $z_2 \leftarrow z_2^{mean} + z_2^{std} \odot \epsilon$ {reparameterisation trick}
29: $m_2 \leftarrow h(s_2)$
30: $y_2 \leftarrow g(s_2, z_2)$
31: $z_{sample2} \leftarrow$ random sample from $\mathcal{N}(\mathbf{0}, \mathbf{I})$
32: $z_{rec2} \leftarrow f_{modality}(g(s_2, z_{sample2}), x_2)$
33:
34: {forward pass for multimodal inputs}
35: $s_1^{deformed} \leftarrow stn(s_1, s_2)$
36: $m_1^{deformed} \leftarrow h(s_1^{deformed})$
37: $y_1^{deformed} \leftarrow g(s_1^{deformed}, z_2)$
38: $s_2^{deformed} \leftarrow stn(s_2, s_1)$
39: $m_2^{deformed} \leftarrow h(s_2^{deformed})$
40: $y_2^{deformed} \leftarrow g(s_2^{deformed}, z_1)$
41:
42: {forward pass for discriminators}
43: $l \leftarrow$ get batch of masks from either modality
44: $m \leftarrow$ sample batch of predicted masks from $\{m_1, m_2, m_1^{deformed}, m_2^{deformed}\}$
45: $rx_1 \leftarrow$ get batch of images from modality 1
46: $px_1 \leftarrow$ sample batch of predicted modality 1 images from $\{g(s_1, z_1), g(s_2^{deformed}, z_1), g(s_1^{deformed}, z_1)\}$
47: $rx_2 \leftarrow$ get batch of images from modality 2
48: $px_2 \leftarrow$ sample batch of predicted modality 2 images from $\{g(s_2, z_2), g(s_2^{deformed}, z_2), g(s_1^{deformed}, z_2)\}$
49: true label $\leftarrow D_M(l)$
50: false label $\leftarrow D_M(m)$
51: true label $\leftarrow D_{I,1}(rx_1)$
52: false label $\leftarrow D_{I,1}(px_1)$
53: true label $\leftarrow D_{I,2}(rx_2)$
54: false label $\leftarrow D_{I,2}(px_2)$
55:
56: {calculate gradients for modality 1 inputs}
57: $G_1 \leftarrow \nabla_{\theta_m, \theta_{a1}} L_{KL}(z_1^{mean}, z_1^{std})$
58: $G_2 \leftarrow \nabla_{\theta_h, \theta_{a1}} L_{sup}(m_1, l_1)$
59: $G_3 \leftarrow -\nabla_{\theta_h, \theta_{a1}} L_{adv}^{M}(m_1)$
60: $G_4 \leftarrow \nabla_{\theta_g, \theta_{a1}, \theta_m} L_{rec}(x_1, y_1)$
61: $G_5 \leftarrow -\nabla_{\theta_g, \theta_{a1}, \theta_m} L_{adv}^{I}(y_1)$
62: $G_6 \leftarrow \nabla_{\theta_g, \theta_m} L_{rec}^{z}(z_{sample1}, z_{rec1})$
63:
64: {calculate gradients for modality 2 inputs}
65: $G_7 \leftarrow \nabla_{\theta_m, \theta_{a2}} L_{KL}(z_2^{mean}, z_2^{std})$
66: $G_8 \leftarrow \nabla_{\theta_h, \theta_{a2}} L_{sup}(m_2, l_2)$
67: $G_9 \leftarrow -\nabla_{\theta_h, \theta_{a2}} L_{adv}^{M}(m_2)$
68: $G_{10} \leftarrow \nabla_{\theta_g, \theta_{a2}, \theta_m} L_{rec}(x_2, y_2)$
69: $G_{11} \leftarrow -\nabla_{\theta_g, \theta_{a2}, \theta_m} L_{adv}^{I}(y_2)$
70: $G_{12} \leftarrow \nabla_{\theta_g, \theta_m} L_{rec}^{z}(z_{sample2}, z_{rec2})$
71:
72: {calculate gradients for multimodal inputs}
73: $G_{11} \leftarrow \nabla_{\theta_h, \theta_{stn}, \theta_{a1} \theta_{a2}} L_{sup}(m_1^{deformed}, l_2)$
74: $G_{12} \leftarrow -\nabla_{\theta_h, \theta_{stn}, \theta_{a1} \theta_{a2}} L_{adv}^{M}(m_1^{deformed})$
75: $G_{13} \leftarrow \nabla_{\theta_h, \theta_{stn}, \theta_{a1} \theta_{a2}} L_{sup}(m_2^{deformed}, l_1)$
76: $G_{14} \leftarrow -\nabla_{\theta_h, \theta_{stn}, \theta_{a1} \theta_{a2}} L_{adv}^{M}(m_2^{deformed})$
77: $G_{15} \leftarrow \nabla_{\theta_g, \theta_{stn}, \theta_{a1} \theta_{a2}, \theta_m} L_{rec}(y_1^{deformed}, x_2)$
78: $G_{16} \leftarrow \nabla_{\theta_g, \theta_{stn}, \theta_{a1} \theta_{a2}, \theta_m} L_{rec}(y_2^{deformed}, x_1)$
79: $G_{17} \leftarrow -\nabla_{\theta_g, \theta_{stn}, \theta_{a1} \theta_{a2}, \theta_m} L_{adv}^{I}(y_1^{deformed})$
80: $G_{18} \leftarrow -\nabla_{\theta_g, \theta_{stn}, \theta_{a1} \theta_{a2}, \theta_m} L_{adv}^{I}(y_2^{deformed})$
81:
82: {accumulate gradients and backpropagate}
83: $\theta_{a1}, \theta_{a2}, \theta_m, \theta_h, \theta_g, \theta_{stn} \leftarrow$ update network parameters using gradients $G_1 \ldots G_{18}$ and Adam optimiser
84:
85: {train discriminators}
86: $G_{19} \leftarrow \nabla_{\psi} L_{adv}^{M}(l, m)$
87: $G_{20} \leftarrow \nabla_{\omega_1} L_{adv}^{I}(rx_1, px_1)$
88: $G_{21} \leftarrow \nabla_{\omega_2} L_{adv}^{I}(rx_2, px_2)$
89: $\psi, \omega_1, \omega_2 \leftarrow$ update network parameters using gradients G_{19}, G_{20}, G_{21} and Adam optimiser
90: **until** convergence of objective