Few-Shot Learning via Saliency-guided Hallucination of Samples (Supplementary Material)

Hongguang Zhang[1,2] Jing Zhang[1,2] Piotr Koniusz[2,1]
[1]Australian National University, [2]Data61/CSIRO
firstname.lastname@{anu.edu.au[1], data61.csiro.au[2]}

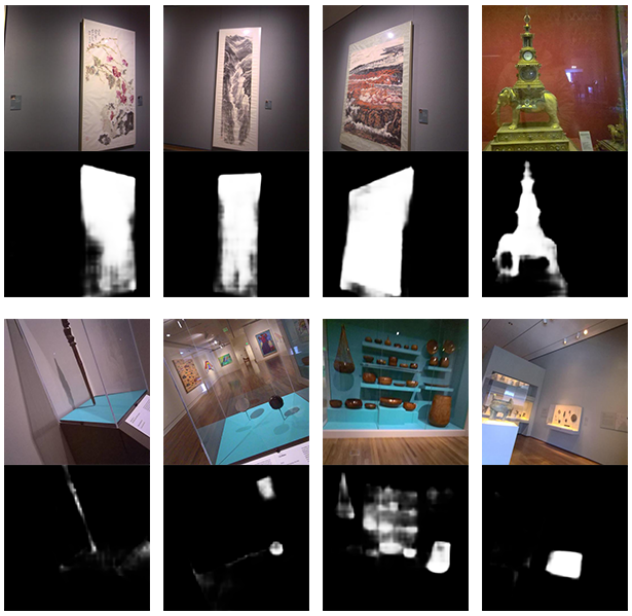

Figure 1: Examples of saliency maps on the Open MIC dataset. The MNL detector was used.

1. Saliency Maps on the Open MIC dataset

In Figure 1, we present saliency maps for some exhibit instances from the Open MIC dataset. Many exhibits can be filtered out reliably. However, saliency maps for composite scenes containing numerous exhibits are the ones most likely to fail. In the future, we will investigate how to improve the use of such unreliable saliency maps for such exhibits. Note that our results on exhibitions containing such composite scences still benefit from our approach–our mixing network can reduce the noise from saliency maps.

2. Evaluations for 224×224 pixel images

We employ 84×84 image in our experiments for fair comparison with other state-of-the-art methods presented in our paper. However, it is easy to use large size images in our network without its modifications due to the ability of second-order representations to aggregate variable number of feature vectors into a fixed-size matrix (our relationship descriptors are stacked matrices). Here we apply 224×224 image to demonstrate the benefits from larger image size.

3. Network Architecture of Our Baseline Models and Additional Experiments for TriR

Below we present the diagrams of two baseline networks used in our paper. The **baseline 1** in Figure 2 is the original pipeline 'w/o Sal. Seg.', which is trained without saliency segmentation or data hallucination – it is very similar to the SoSN pipeline [7]. Figure 3 demonstrates the **baseline 2** 'w/o Hal.', which employs saliency network to segment the foregrounds and backgrounds but does not hallucinate the data (no mixing of a foreground with numerous different backgrounds is allowed).

In our paper, the reported results are obtained by using **baseline 2** 'w/o Hal.' pipeline as teacher in TriR regulariza-

Table 1: Accuracy on the *mini*Imagenet dataset given different size of images. See [5, 7] for details of baselines. The astersik (*) denotes the 'sanity check' results on our proposed pipeline given disabled both saliency segmentation and hallucination.

Model	Fine Tune	5-way Acc. 1-shot	5-shot
Matching Nets [6]	N	43.56 ± 0.84	55.31 ± 0.73
Meta Nets [2]	N	49.21 ± 0.96	-
Meta-Learn Nets [3]	N	43.44 ± 0.77	60.60 ± 0.71
Prototypical Net [4]	N	49.42 ± 0.78	68.20 ± 0.66
MAML [1]	Y	48.70 ± 1.84	63.11 ± 0.92
Relation Net [5]	N	51.36 ± 0.86	65.63 ± 0.72
SoSN [7]	N	52.96 ± 0.83	68.63 ± 0.68
84×84			
SalNet w/o Sal. Seg. (*)	N	53.15 ± 0.87	68.87 ± 0.67
SalNet w/o Hal.	N	55.57 ± 0.86	70.35 ± 0.66
SalNet Intra-class Hal.	N		71.78 ± 0.69
SalNet Inter-class Hal.	N	57.45 ± 0.88	72.01 ± 0.67
224×224			
SoSN[7]	N	59.22 ± 0.91	73.24 ± 0.69
SalNet w/o Sal. Seg. (*)	N	60.36 ± 0.86	74.34 ± 0.67
SalNet w/o Hal.	N	62.22 ± 0.87	76.86 ± 0.65
SalNet Intra-class Hal.	N		77.95 ± 0.65
SalNet Inter-class Hal.	N	63.88 ± 0.86	78.34 ± 0.63

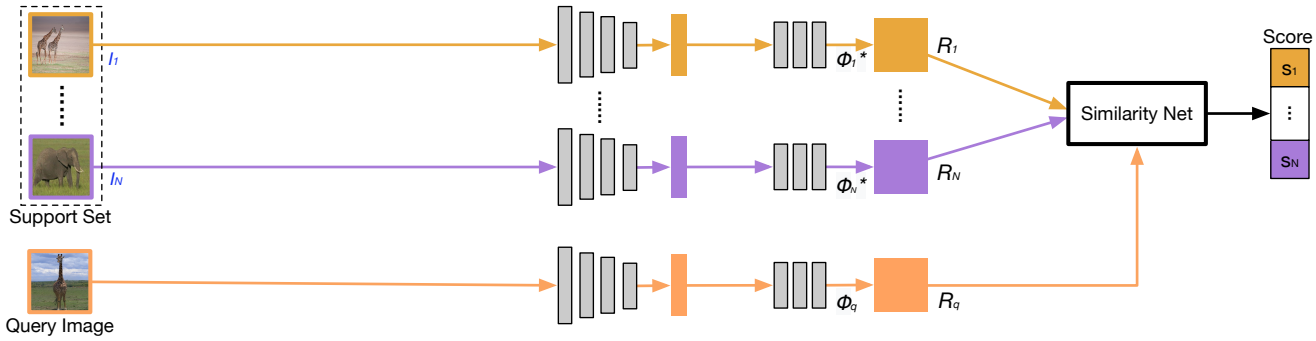

Figure 2: The network architecture of **baseline 1** 'w/o Sal. Seg.'. It can be seen that once the Saliency Net and data hallucination strategies are disabled, the network pipeline are very similar to SoSN . Note that we write $\mathbf{R}_1, ..., \mathbf{R}_N$ for brevity rather than $\mathbf{R}_{11}, ..., \mathbf{R}_{NN}$ (as dictated by Eq. (4) of our main submission) as no hallucination takes place here *e.g.*, we evaluate only $\mathbf{R}_{ij}$ for $i = j$. Moreover, note that Φ_i are not generated by the foreground-background mechanism from Eq. (3) of our main submission. Instead, entire images are encoded.

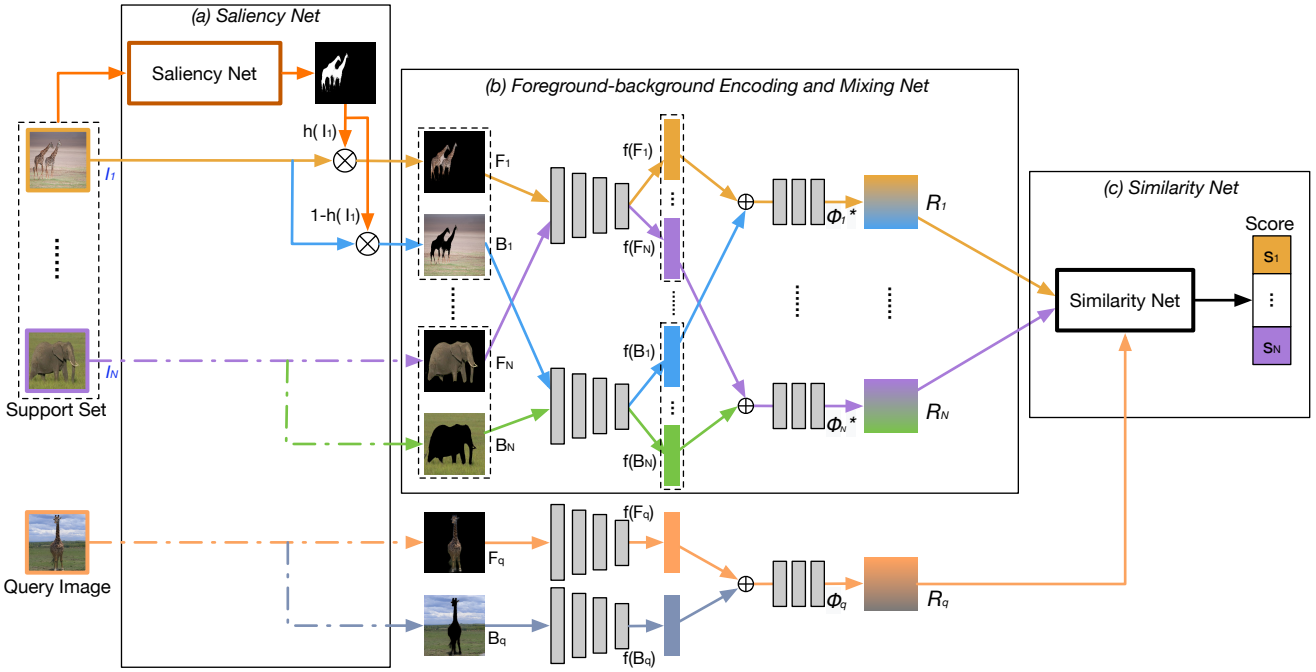

Figure 3: The network architecture of **baseline 2** 'w/o Hallucination'. Note that although the data hallucination mechanism is disabled, we still apply saliency maps to segment foregrounds and backgrounds as we want the TriR loss to learn to account for the potential noise stemming from the foreground-background segmentation which is used in the main network. Moreover, we write $\mathbf{R}_1, ..., \mathbf{R}_N$ for brevity rather than $\mathbf{R}_{11}, ..., \mathbf{R}_{NN}$ (as dictated by Eq. (4) of our main submission) as no hallucination takes place here *e.g.*, we evaluate only $\mathbf{R}_{ij}$ for $i = j$. Note that Φ_i are generated by the foreground-background mechanism in Eq. (3) of our main paper (we abbreviate Φ_{ii} to Φ_i).

Table 2: Evaluations on the *mini*Imagenet dataset given different teacher networks for the TriR regularization.

Model	Fine Tune	5-way Acc. 1-shot	5-shot
baseline 1 (opt. (i)) as a teacher network in TriR			
SalNet Intra-class Hal.	N	56.11 ± 0.88	71.56 ± 0.67
SalNet Inter-class Hal.	N	57.24 ± 0.94	72.49 ± 0.65
baseline 2 (opt. (ii)) as a teacher network in TriR			
SalNet Intra-class Hal.	N	55.57 ± 0.86	71.78 ± 0.69
SalNet Inter-class Hal.	N	57.45 ± 0.88	72.01 ± 0.67

tion (option (ii) in line 513 of our main submission). However, for completeness, we also investigate **baseline 1** 'w/o Hal.' pipeline as teacher in TriR regularization (option (i) in line 512 of our main submission). Table 2 shows that both TriR teachers perform similarly to each other.

References

[1] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, pages 1126–1135, 2017. 1

[2] T. Munkhdalai and H. Yu. Meta networks. *CoRR:1703.00837*,

2017. 1

[3] S. Ravi and H. Larochelle. Optimization as a model for few-shot learning. In *ICLR*, 2017. 1

[4] J. Snell, K. Swersky, and R. Zemel. Prototypical networks for few-shot learning. In *NIPS*, pages 4077–4087, 2017. 1

[5] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. Learning to compare: Relation network for few-shot learning. *CoRR:1711.06025*, 2017. 1

[6] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al. Matching networks for one shot learning. In *NIPS*, pages 3630–3638, 2016. 1

[7] H. Zhang and P. Koniusz. Power normalizing second-order similarity network for few-shot learning. In *WACV*, pages 1185–1193. IEEE, 2019. 1