Supplementary File for *n*-Reference Transfer Learning for Saliency Prediction

Yan Luo[1 [0000−0001−5135−0316]], Yongkang Wong[2 [0000−0002−1239−4428]], Mohan S. Kankanhalli[2 [0000−0002−4846−2015]], and Qi Zhao[1 [0000−0003−3054−8934]]

[1] Department of Computer Science and Engineering, University of Minnesota
luoxx648@umn.edu, qzhao@cs.umn.edu
[2] School of Computing, National University of Singapore
{wongyk, mohan}@comp.nus.edu.sg

In this supplementary document, we first present the proof of Theorem 1. Then, we provide both the mean and standard deviation of scores on NSS, AUC, and CC metrics within the four settings, i.e., ⟨SALICON, DINet, WebSal⟩ (see Table 1), ⟨SALICON, ResNet-50, WebSal⟩ (see Table 2), ⟨MIT1003, DINet, WebSal⟩ (see Table 3), and ⟨SALICON, DINet, Art⟩ (see Table 4). Overall, the standard deviations generated by the proposed TR−Ref and FT|Ref are similar to the ones generated by TR and FT, respectively.

Theorem 1 (Saliency generalization bound). *Denote H as a finite hypothesis set. Given ℓ^p and $y \in [0,1]^m$, for any $\delta > 0$, with probability at least $1-\delta$, the following inequality holds for all $f \in H$:*

$$|R_{\mathcal{D}}(f) - \hat{R}_{\mathcal{D}}(f)| \leq m^{\frac{1}{p}} \sqrt{\frac{\log |H| + \log \frac{2}{\delta}}{2|D|}}$$

Proof. The proof sketch is similar to the regression generalization bound provided in [1]. First, as $\ell^p(y_1, y_2) = (\sum_i^m |y_{1i} - y_{2i}|^p)^{\frac{1}{p}} \leq m^{\frac{1}{p}}$, we know ℓ^p is bounded by $m^{\frac{1}{p}}$. Then, by the union bound, given an error ξ, we have

$$Pr[\sup_{f \in H} |R(f) - \hat{R}(f)| > \xi] \leq \sum_{f \in H} Pr[|R(f) - \hat{R}(f)| > \xi].$$

By Hoeffding's bound, we have

$$\sum_{f \in H} Pr[|R(f) - \hat{R}(f)| > \xi] \leq 2|H| \exp(-\frac{2|D|\xi^2}{m^{\frac{2}{p}}}).$$

Due to the probability definition, $2|H|\exp(-\frac{2|D|\xi^2}{m^{\frac{2}{p}}}) = \delta$. Considering ξ is a function of other variables, we can rearrange it as $\xi = m^{\frac{1}{p}} \sqrt{\frac{\log|H| + \log\frac{2}{\delta}}{2|D|}}$. Since we know $Pr[|R(f) - \hat{R}(f)| > \xi]$ is with probability at most δ, it can be inferred that $Pr[|R(f) - \hat{R}(f)| <= \xi]$ is at least $1-\delta$. □

Table 1: Performance within setting ⟨SALICON, DINet, WebSal⟩. ↑ implies that higher score is better. The score in bold font indicates the best result under the metric. We take 10 runs for conventional training, 1-shot, 5-shot, and 10-shot transfer learning and report the mean and the std. Empirical upper bound (EUB) is generated by the 3-fold cross validation on the target domain dataset. The details of the experimental setup are provided in Section 4.1

		NSS↑	AUC↑	CC↑
TR	$n = 0$	1.3330±0.0084	0.7796±0.0025	0.5515±0.0033
TR−Ref	$n = 1$	1.3621±0.0191	0.7848±0.0023	0.5628±0.0073
FT	$n = 1$	1.4731±0.0466	0.8005±0.0120	0.5976±0.0175
FT|Ref	$n = 1$	**1.5077**±0.0497	**0.8051**±0.0121	**0.6121**±0.0171
TR−Ref	$n = 5$	1.3683±0.0266	0.7874±0.0083	0.5659±0.0124
FT	$n = 5$	1.5803±0.0346	0.8161±0.0062	0.6355±0.0113
FT|Ref	$n = 5$	**1.6085**±0.0212	**0.8200**±0.0050	**0.6468**±0.0084
TR−Ref	$n = 10$	1.3647±0.0150	0.7839±0.0046	0.5633±0.0065
FT	$n = 10$	1.6290±0.0214	0.8247±0.0048	0.6531±0.0085
FT|Ref	$n = 10$	**1.6439**±0.0249	**0.8276**±0.0056	**0.6605**±0.0095
TR−Ref	EUB	1.3822±0.0413	0.7910±0.0159	0.5708±0.0229
FT	EUB	1.8695±0.0268	0.8488±0.0051	0.7389±0.0047
FT|Ref	EUB	**1.8831**±0.0189	**0.8494**±0.0046	**0.7442**±0.0058

Table 2: Performance within setting ⟨SALICON, ResNet-50, WebSal⟩

		NSS↑	AUC↑	CC↑
TR	$n = 0$	1.2950±0.0604	0.7749±0.0120	0.5358±0.0226
TR−Ref	$n = 1$	1.3569±0.0236	0.7864±0.0055	0.5611±0.0099
FT	$n = 1$	1.3722±0.0611	0.7923±0.0169	0.5627±0.0254
FT|Ref	$n = 1$	**1.4272**±0.0565	**0.7983**±0.0133	**0.5817**±0.0224
TR−Ref	$n = 5$	1.3535±0.0159	0.7837±0.0080	0.5593±0.0078
FT	$n = 5$	1.5043±0.0165	0.8131±0.0051	0.6139±0.0064
FT|Ref	$n = 5$	**1.5491**±0.0242	**0.8149**±0.0083	**0.6281**±0.0092
TR−Ref	$n = 10$	1.3583±0.0242	0.7857±0.0056	0.5612±0.0103
FT	$n = 10$	1.5164±0.0224	0.8103±0.0051	0.6200±0.0078
FT|Ref	$n = 10$	**1.5829**±0.0282	**0.8143**±0.0082	**0.6414**±0.0100
TR−Ref	EUB	1.3626±0.0082	0.7864±0.0110	0.5645±0.0031
FT	EUB	1.8325±0.0414	0.8462±0.0016	0.7275±0.0066
FT|Ref	EUB	**1.8500**±0.0391	**0.8480**±0.0016	**0.7321**±0.0049

References

1. Mohri, M., Rostamizadeh, A., Talwalkar, A.: Foundations of Machine Learning. The MIT Press (2012)

Table 3: Performance within setting ⟨MIT1003, DINet, WebSal⟩

		NSS↑	AUC↑	CC↑
TR	$n = 0$	1.3905±0.0062	0.7991±0.0010	0.5700±0.0021
TR−Ref	$n = 1$	1.4405±0.0175	0.8085±0.0026	0.5902±0.0071
FT	$n = 1$	1.4410±0.0941	0.8023±0.0148	0.5784±0.0397
FT|Ref	$n = 1$	**1.4575**±0.0884	**0.8070**±0.0140	**0.5838**±0.0369
TR−Ref	$n = 5$	1.4452±0.0156	0.8064±0.0025	0.5908±0.0055
FT	$n = 5$	1.5795±0.0322	0.8217±0.0058	0.6395±0.0134
FT|Ref	$n = 5$	**1.6136**±0.0304	**0.8269**±0.0045	**0.6515**±0.0116
TR−Ref	$n = 10$	1.4330±0.0165	0.8060±0.0033	0.5872±0.0055
FT	$n = 10$	1.6462±0.0216	0.8261±0.0044	0.6660±0.0083
FT|Ref	$n = 10$	**1.6691**±0.0206	**0.8283**±0.0039	**0.6730**±0.0068
TR−Ref	EUB	1.4402±0.0410	0.8087±0.0035	0.5905±0.0108
FT	EUB	1.8450±0.0430	0.8466±0.0036	0.7330±0.0086
FT|Ref	EUB	**1.8507**±0.0397	**0.8478**±0.0038	**0.7344**±0.0080

Table 4: Performance within setting ⟨SALICON, DINet, Art⟩

		NSS↑	AUC↑	CC↑
TR	$n = 0$	1.5172±0.0525	0.8225±0.0063	0.6003±0.0184
TR−Ref	$n = 1$	1.5651±0.0299	0.8287±0.0045	0.6211±0.0105
FT	$n = 1$	1.6255±0.0365	0.8324±0.0095	0.6449±0.0153
FT|Ref	$n = 1$	**1.6523**±0.0397	**0.8380**±0.0088	**0.6564**±0.0150
TR−Ref	$n = 5$	1.5870±0.0238	0.8304±0.0029	0.6274±0.0093
FT	$n = 5$	1.8049±0.0317	0.8480±0.0056	0.7185±0.0161
FT|Ref	$n = 5$	**1.8314**±0.0341	**0.8503**±0.0051	**0.7274**±0.0152
TR−Ref	$n = 10$	1.5704±0.0377	0.8288±0.0036	0.6204±0.0139
FT	$n = 10$	1.8325±0.0819	0.8474±0.0050	0.7288±0.0301
FT|Ref	$n = 10$	**1.8584**±0.0949	**0.8503**±0.0043	**0.7366**±0.0341
TR−Ref	EUB	1.5980±0.0347	0.8340±0.0043	0.6331±0.0152
FT	EUB	2.1595±0.0764	0.8636±0.0063	0.8464±0.0054
FT|Ref	EUB	**2.1874**±0.0795	**0.8649**±0.0065	**0.8519**±0.0110