A General Divergence Modeling Strategy for Salient Object Detection

(Supplementary Materials)

Xinyu Tian[1], Jing Zhang[2], and Yuchao Dai[1]✉

[1] Northwestern Polytechnical University, Xi'an, China
[2] Australian National University, Canberra, Australia

1 Evaluation Metrics

We evaluate the performance of ours and competing methods with three saliency evaluation metrics, including: Mean Absolute Error ($\mathcal{M}$), mean F-measure (F_β) and mean E-measure (E_ξ) [1] . We also present an uncertainty based mean absolute error to estimate the divergence modeling ability of each model.

MAE $\mathcal{M}$ is defined as the per-pixel wise difference between predicted saliency map s and a per-pixel wise binary ground-truth y:

$$\text{MAE} = \frac{1}{H \times W}|s - y|, \tag{1}$$

where H and W are height and width of s. MAE provides a direct estimate of conformity between estimated and ground-truth maps. However, for the MAE metric, small objects naturally assign a smaller error and larger objects have larger errors.

F-measure F_β is a region based similarity metric, and we provide the mean F-measure using varying fixed (0-255) thresholds.

E-measure E_ξ is the recent proposed Enhanced alignment measure [1] in the binary map evaluation field to jointly capture image-level statistics and local pixel matching information.

Uncertainty Based MAE: The proposed uncertainty based MAE is used to evaluate the accuracy of the generated uncertainty map. Given multiple annotations $\{y^j\}_{j=1}^M$, we compute its mean annotation $\bar{y} = \frac{1}{M}\sum_{j=1}^M y^j$. The ground truth uncertainty is then defined as the entropy of the mean annotation: $U_{gt} = \mathbb{H}[\bar{y}]$. At test time, we run multiple times of sampling (for the latent variable models) or directly define the predictions from the multi-head structure of deep ensemble as stochastic predictions. We then obtain the mean prediction and its entropy (uncertainty) of each stochastic model. The uncertainty based MAE is defined as the MAE of the predicted uncertainty and the ground truth uncertainty U_{gt}.

Yuchao Dai (daiyuchao@nwpu.edu.cn) is the corresponding author. This work was supported in part by the NSFC (61871325). Code is available at https://npucvr.github.io/Divergence_SOD/.

Table 1. Performance of the proposed strategy within the ensemble based frameworks.

	DUTS [2]			ECSSD [3]			DUT [4]			HKU-IS [5]			COME-E [6]			COME-H [6]		
	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$
Base_A	.811	.897	.045	.903	.931	.044	.741	.846	.059	.887	.935	.036	.852	.905	.049	.807	.859	.080
Base_M	.824	.911	.041	.916	.946	.038	.754	.864	.055	.900	.950	.032	.877	.927	.040	.838	.882	.070
DE_A	.819	.906	.042	.910	.939	.040	.744	.855	.057	.895	.940	.034	.870	.920	.043	.831	.876	.073
DE_R	.816	.900	.043	.916	.944	.038	.738	.843	.060	.892	.939	.036	.869	.920	.043	.832	.880	.072

Table 2. Performance of the proposed strategy within the GAN based frameworks.

	DUTS [2]			ECSSD [3]			DUT [4]			HKU-IS [5]			COME-E [6]			COME-H [6]		
	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$
Base_A	.811	.897	.045	.903	.931	.044	.741	.846	.059	.887	.935	.036	.852	.905	.049	.807	.859	.080
Base_M	.824	.911	.041	.916	.946	.038	.754	.864	.055	.900	.950	.032	.877	.927	.040	.838	.882	.070
GAN_A	.803	.895	.045	.906	.939	.040	.736	.850	.060	.879	.931	.038	.855	.912	.046	.816	.867	.076
GAN_R	.817	.906	.041	.914	.944	.039	.741	.849	.059	.895	.946	.033	.873	.926	.040	.834	.882	.070

Table 3. Performance of the proposed strategy within the VAE based frameworks.

	DUTS [2]			ECSSD [3]			DUT [4]			HKU-IS [5]			COME-E [6]			COME-H [6]		
	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$
Base_A	.811	.897	.045	.903	.931	.044	.741	.846	.059	.887	.935	.036	.852	.905	.049	.807	.859	.080
Base_M	.824	.911	.041	.916	.946	.038	.754	.864	.055	.900	.950	.032	.877	.927	.040	.838	.882	.070
VAE_A	.798	.888	.049	.910	.940	.039	.727	.843	.068	.890	.937	.036	.853	.909	.048	.813	.865	.078
VAE_R	.822	.899	.042	.913	.936	.042	.723	.824	.062	.895	.937	.036	.880	.924	.041	.838	.879	.071

Table 4. Performance of the proposed strategy within the ABP based frameworks.

	DUTS [2]			ECSSD [3]			DUT [4]			HKU-IS [5]			COME-E [6]			COME-H [6]		
	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$	$F_\beta\uparrow$	$E_\xi\uparrow$	$\mathcal{M}\downarrow$
Base_A	.811	.897	.045	.903	.931	.044	.741	.846	.059	.887	.935	.036	.852	.905	.049	.807	.859	.080
Base_M	.824	.911	.041	.916	.946	.038	.754	.864	.055	.900	.950	.032	.877	.927	.040	.838	.882	.070
ABP_A	.777	.866	.052	.888	.916	.049	.720	.829	.063	.853	.900	.046	.835	.886	.054	.794	.844	.085
ABP_R	.805	.896	.047	.911	.938	.044	.734	.846	.060	.885	.937	.039	.864	.911	.050	.824	.863	.083

2 Performance Comparison of Divergence Modeling Techniques

We provide more comparisons of the divergence modeling techniques in this section. Note that, the baseline models training with all pairs of samples ($D2 = \{\{x_i^j, y_i^j\}_{j=0}^M\}_{i=1}^N$) and training with only the majority voting ground truth ($D0 = \{x_i^j, y_i^0\}_{i=1}^N$) are denoted as "Base_A" and "Base_M" respectively.

2.1 Deep Ensemble

We show deterministic prediction of the deep ensemble framework in Table 1, where "DE_R" is trained with the proposed random sampling strategy, and "DE_A" is trained with $D2$. The comparable performance of "DE_A" and "DE_R" explains that our random sampling strategy will not decrease the performance. Then, we show the produced uncertainty maps in Fig. 1, which further explains the superiority of our strategy in saliency divergence modeling.

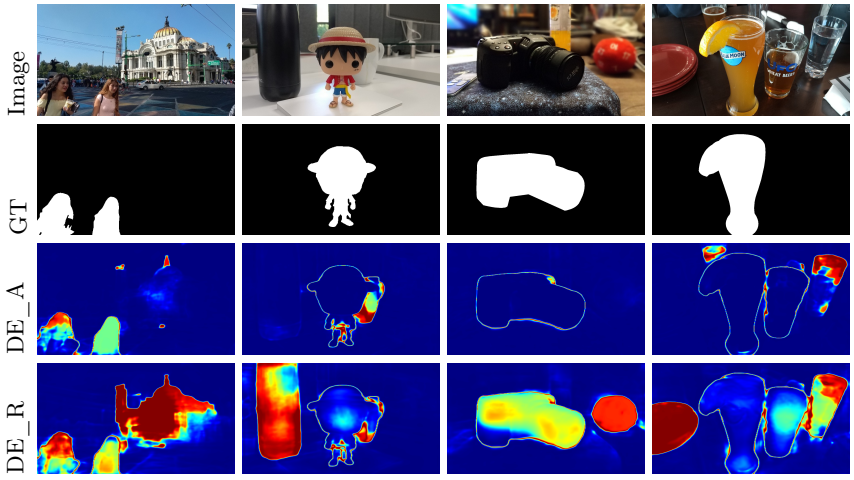

Fig. 1. Uncertainty comparison of the deep ensemble framework with ("DE_A")/without ("DE_R") the proposed random sampling strategy.

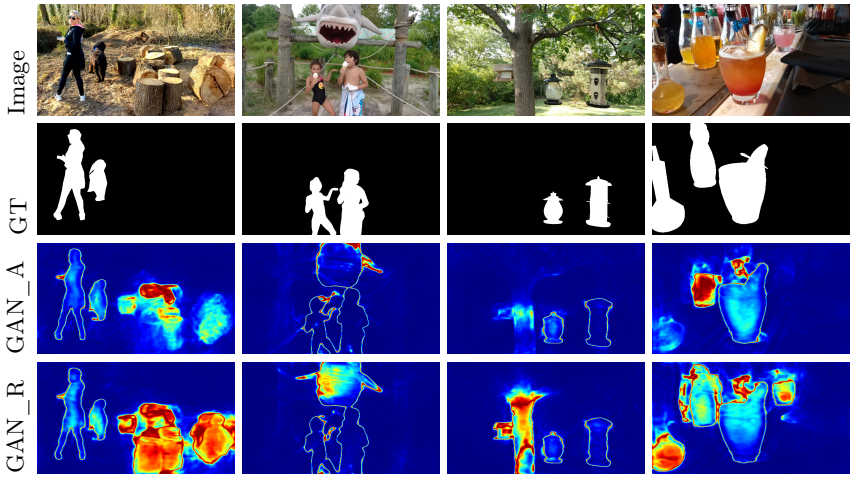

Fig. 2. Uncertainty comparison of the GAN framework with ("GAN_A")/without ("GAN_R") the proposed random sampling strategy.

2.2 Latent Variable Models

Similarly, we show model performance of each latent variable model training with all annotations (D_2) and our random sampling strategy in Table 2, 3 and 4, indicating the GAN [7], VAE [8] and ABP [9] based latent variable models respectively. Again, the comparable deterministic performance of each latent

Fig. 3. Uncertainty comparison of the VAE framework with ("VAE_A")/without ("VAE_R") the proposed random sampling strategy.

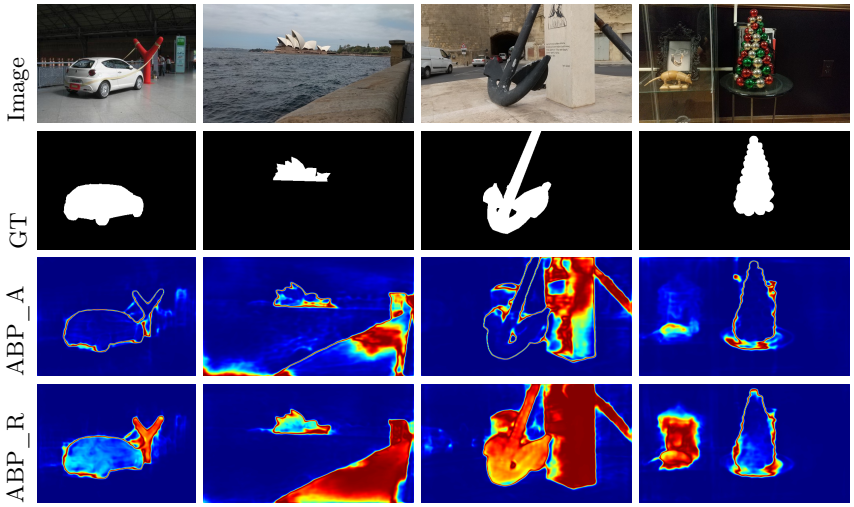

Fig. 4. Uncertainty comparison of the ABP framework with ("ABP_A")/without ("ABP_R") the proposed random sampling strategy.

variable model with/without our random sampling strategy shows that our random sampling strategy can maintain the model deterministic performance.

We further show the generated uncertainty maps of each method with/without our random sampling strategy in Fig. 2, 3 and 4 to explain the superiorty of the proposed random sampling strategy.

Table 5. Performance comparison with model trained with majority voting ground truth within the GAN based frameworks.

	DUTS [2]			ECSSD [3]			DUT [4]			HKU-IS [5]			COME-E [6]			COME-H [6]		
	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$
Base_M	.824	.911	.041	.916	.946	.038	.754	.864	.055	.900	.950	.032	.877	.927	.040	.838	.882	.070
GAN_M	.817	.905	.042	.913	.945	.038	.747	.855	.058	.893	.944	.035	.870	.923	.042	.831	.881	.071
GAN_R	.817	.906	.041	.914	.944	.039	.741	.849	.059	.895	.946	.033	.873	.926	.040	.834	.882	.070

Table 6. Performance comparison with model trained with majority voting ground truth within the VAE based frameworks.

	DUTS [2]			ECSSD [3]			DUT [4]			HKU-IS [5]			COME-E [6]			COME-H [6]		
	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$
Base_M	.824	.911	.041	.916	.946	.038	.754	.864	.055	.900	.950	.032	.877	.927	.040	.838	.882	.070
VAE_M	.817	.905	.042	.913	.945	.038	.744	.852	.058	.893	.943	.035	.873	.924	.042	.831	.880	.072
VAE_R	.822	.899	.042	.913	.936	.042	.723	.824	.062	.895	.937	.036	.880	.924	.041	.838	.879	.071

Table 7. Performance comparison with model trained with majority voting ground truth within the ABP based frameworks.

	DUTS [2]			ECSSD [3]			DUT [4]			HKU-IS [5]			COME-E [6]			COME-H [6]		
	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$	$F_\beta \uparrow$	$E_\xi \uparrow$	$\mathcal{M} \downarrow$
Base_M	.824	.911	.041	.916	.946	.038	.754	.864	.055	.900	.950	.032	.877	.927	.040	.838	.882	.070
ABP_M	.801	.893	.045	.908	.942	.040	.732	.840	.060	.886	.937	.038	.862	.916	.045	.825	.874	.074
ABP_R	.805	.896	.047	.911	.938	.044	.734	.846	.060	.885	.937	.039	.864	.911	.050	.824	.863	.083

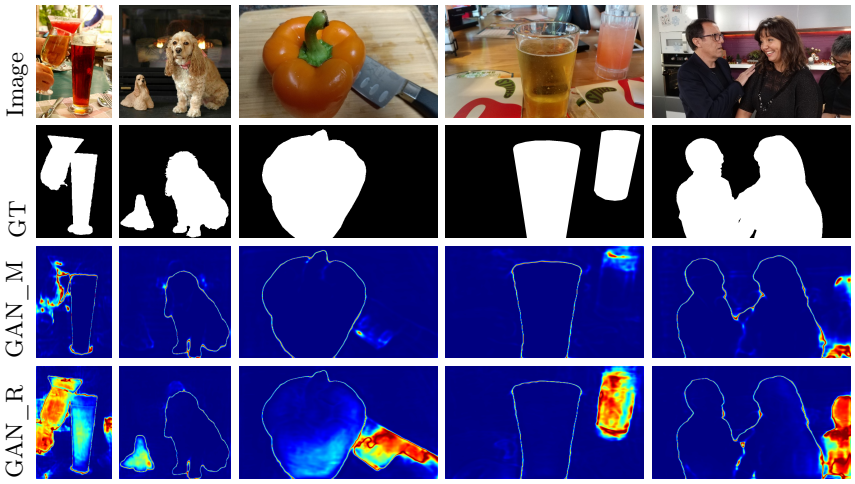

Fig. 5. Uncertainty comparison of the GAN framework with majority voting annotation and with multiple annotations by using the proposed random sampling strategy.

3 Uncertainty with Majority Voting GT

In this section, we further compare divergence modeling performance of generative models with majority voting ground truth y^0 and with multiple annotations

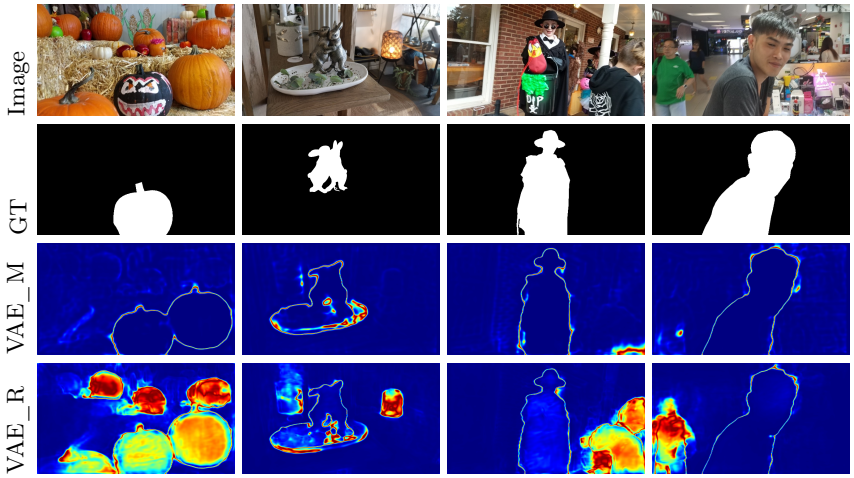

Fig. 6. Uncertainty comparison of the VAE framework with majority voting annotation and with multiple annotations by using the proposed random sampling strategy.

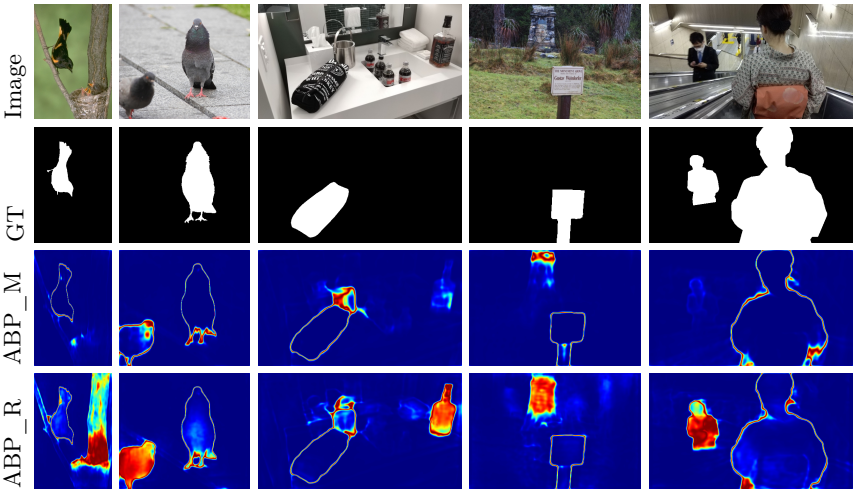

Fig. 7. Uncertainty comparison of the ABP framework with majority voting annotation and with multiple annotations by using the proposed random sampling strategy.

$\{y^j\}_{j=1}^M$ by using the proposed random sampling strategy. The deterministic performance is shown in Table 5, 6 and 7, representing the GAN, VAE and ABP based frameworks respectively. The comparable deterministic performance indicates again that our random sampling strategy will not decrease model deterministic predictions.

Table 8. Performance of applying the proposed random sampling strategy to two other SOTA SOD models. "-E", "-G", "-V" and "-A" indicate the corresponding deep ensemble, GAN, VAE and ABP based model.

	DUTS			ECSSD			DUT			HKU-IS			COME-E			COME-H		
Method	F_β	E_ξ	$\mathcal{M}$	F_β	E_ξ	$\mathcal{M}$	F_β	E_ξ	$\mathcal{M}$	F_β	E_ξ	$\mathcal{M}$	F_β	E_ξ	$\mathcal{M}$	F_β	E_ξ	$\mathcal{M}$
MINet	.820	.902	.041	.920	.947	.037	.751	.852	.053	.898	.945	.032	.877	.924	.040	.840	.884	.068
-E	.835	.912	.040	.926	.951	.034	.763	.864	.056	.903	.947	.032	.888	.931	.037	.849	.889	.065
-G	.839	.914	.038	.927	.950	.034	.768	.870	.052	.903	.946	.032	.891	.933	.037	.853	.890	.065
-V	.830	.907	.041	.923	.946	.036	.749	.866	.056	.897	.940	.034	.891	.931	.037	.849	.886	.067
-A	.828	.906	.040	.921	.945	.037	.747	.849	.055	.896	.938	.035	.887	.928	.039	.842	.878	.072
GateNet	.820	.898	.042	.920	.943	.038	.749	.847	.056	.897	.939	.035	.876	.922	.041	.839	.880	.070
-E	.830	.905	.044	.925	.949	.036	.760	.861	.063	.898	.941	.036	.882	.925	.041	.842	.881	.072
-G	.850	.922	.036	.926	.950	.035	.782	.877	.050	.908	.949	.031	.892	.933	.037	.852	.891	.065
-V	.845	.917	.038	.928	.950	.035	.775	.871	.052	.905	.946	.032	.892	.932	.037	.852	.889	.066
-A	.840	.915	.039	.922	.947	.037	.772	.871	.055	.904	.947	.033	.886	.929	.040	.845	.884	.070

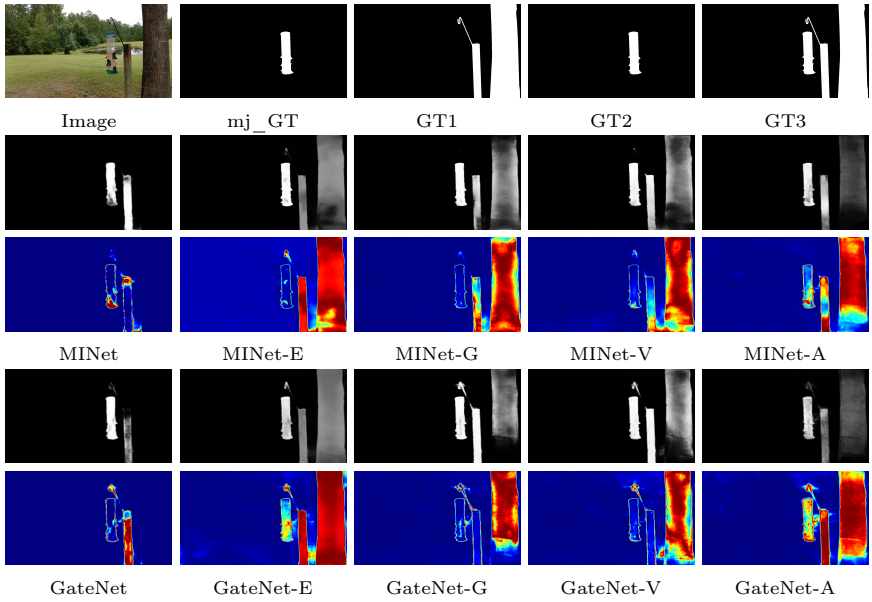

Fig. 8. Uncertainty maps of two SOTA SOD model (MINet[10] and GateNet[11]) with the proposed general divergence modeling strategy. The first column shows image, the majority voting ground truth and labels from different annotators. From the second column to the last one, we show prediction (top) and the corresponding predictive uncertainty (bottom).

We further show the generated uncertainty maps of each latent variable model with majority voting annotation and multiple annotations using the proposed strategy, and the results are shown in Fig.5, 6 and 7. The more uniformly activated uncertainty regions further explain the superior performance of our proposed strategy in saliency divergence modeling.

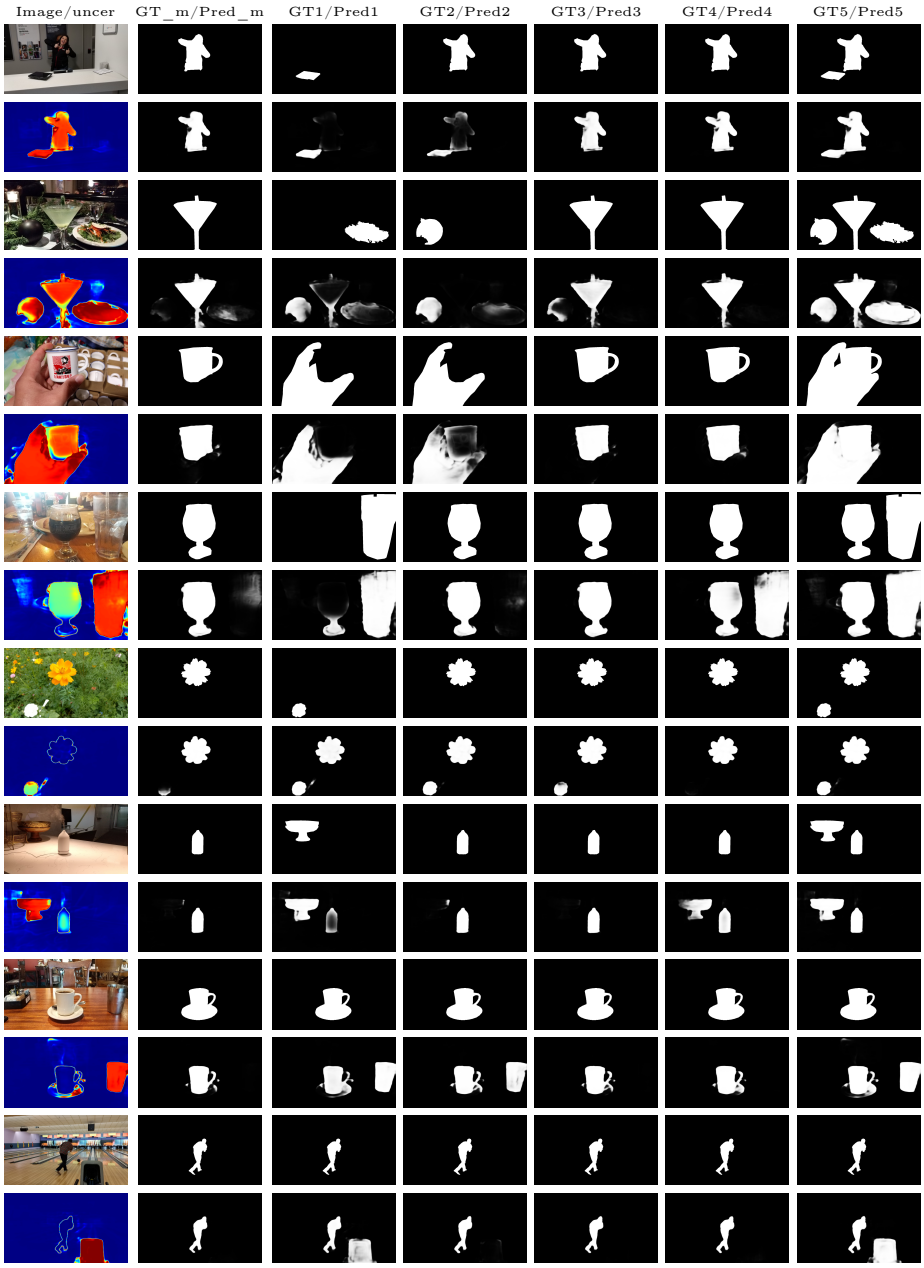

Fig. 9. Generated diverse saliency maps. For each example, the first column shows, from left to right, input image, ground truth map after majority voting, and ground truth maps from 5 different annotators; the second column shows the predicted uncertainty map, prediction of our majority voting branch, and five generated diverse saliency maps.

4 Applying Our Solution to Other SOTA SOD Models

We additionally apply our strategy to two other SOTA SOD models, namely MINet[10] and GateNet[11]. The deterministic performance is reported in Table 8, which clear shows that our random sampling strategy can in general keep the deterministic performance untouched. We further visualize the produced uncertainty maps of our methods in Fig. 8. The generated uncertainty maps of our methods can indeed explain the less salient regions, which verifies the superiority of our proposed strategy.

5 Generated saliency maps

In addition to the uncertainty maps shown previously, we also show our generated diverse saliency maps in Fig. 9, where each pair (top-down) of images are explained in the corresponding text on top of the column. *e.g.*. "Image/uncer" indicates the image/uncertainty map pair; "GT_m/Pred_m" is the pair showing the majority ground truth and the prediction from our majority prediction branch; "GT**M**/Pred**M**" (M=1,2,3,4,5) represents the pair of diverse annotation and the corresponding prediction. Note that, we pair the prediction to the diverse annotation via nearest neighbor searching only for visualization. In practice, we do not pair them. Fig. 9 further demonstrate the superiority of our divergence modeling strategy. Specifically, the last two examples in Fig. 9 reveal that our models have the potential to discover the less salient objects that are not presented in the multiple ground truth annotations.

References

1. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment measure for binary foreground map evaluation. International Joint Conference on Artificial Intelligence (IJCAI) (2018) 698–704
2. Wang, L., Lu, H., Wang, Y., Feng, M., Wang, D., Yin, B., Ruan, X.: Learning to detect salient objects with image-level supervision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017) 136–145
3. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2013) 1155–1162
4. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2013) 3166–3173
5. Li, G., Yu, Y.: Visual saliency based on multiscale deep features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015) 5455–5463
6. Zhang, J., Fan, D.P., Dai, Y., Yu, X., Zhong, Y., Barnes, N., Shao, L.: Rgb-d saliency detection via cascaded mutual information minimization. In: IEEE International Conference on Computer Vision (ICCV). (2021) 4338–4347
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Conference on Neural Information Processing Systems (NeurIPS). Volume 27. (2014) 2672–2680

8. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (ICLR). (2014)
9. Han, T., Lu, Y., Zhu, S.C., Wu, Y.N.: Alternating back-propagation for generator network. In: AAAI Conference on Artificial Intelligence (AAAI). (2017)
10. Pang, Y., Zhao, X., Zhang, L., Lu, H.: Multi-scale interactive network for salient object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2020) 9413–9422
11. Zhao, X., Pang, Y., Zhang, L., Lu, H., Zhang, L.: Suppress and balance: A simple gated network for salient object detection. In: European Conference on Computer Vision (ECCV). (2020) 35–51